

Congressional agendas

A US Congress newly dominated by Democrats needs to exercise financial restraint. Manned spaceflight is a good candidate for cuts, and energy research needs belated leadership.

The Democrats who will make up majorities in both houses of the 110th US Congress in January face a difficult two years. Despite having their hands once again on the purse strings, without control of the executive there will be many areas where their ability to make good on the change they have promised will be constrained. But there is good that can be done on various issues of particular interest to the research community.

One tempting possibility is to use Congress's investigative powers to delve into the Bush administration's past and future use, and alleged abuse, of scientific advice. That may be useful in setting standards for future administrations of either party, and might well have practical political benefits. But it is not, for now, the highest priority.

That honour goes to policy on energy and climate. Congress will doubtless be keen to assert US leadership in this area; climate is, among other things, a particular interest of Senator Joe Biden, the likely chair of the Senate Committee on Foreign Relations. And the world would undoubtedly benefit from wise leadership over the next two years as the post-Kyoto agenda becomes clear. But any leadership worthy of the name must be by example.

There are various proposals for cap and trade approaches to greenhouse gases that Democratic majorities might introduce to Congress — notably a bill sponsored by senators John McCain (Republican, Arizona) and Joe Lieberman (Democrat, Connecticut). Unfortunately, with the likelihood of vigorous opposition from the oil and gas lobby and the current administration's firm position on the other side of the argument, passing such legislation, while desirable, might not be possible. If that proves to be the case, it will be time to abandon the aggrandizing rhetoric of leadership, and to settle for signalling a statesmanlike interest in 'followership'. Finding ways to demonstrate that the United States is coming into line with informed opinion and policy in the rest of the world would in itself be a step forward.

Meanwhile, a strengthened commitment to energy research and

development, and an honest approach to the evidence base in this area, should be more achievable. Congress should not turn its back on the possibility of renewed investment in nuclear power, as some Democrats will be keen to do. But it should also set up a proactive congressional review of which other energy technologies are ripe for development and innovation. Energy independence is a popular theme in the United States; solutions to the problem that also reduce carbon emissions have a clear mandate.

The new congressional leaders will find themselves in a more difficult position on the scientific issue that had the highest profile in the campaign: stem-cell research. The stance that led to President George W. Bush's regrettable veto of legislation that would have relaxed constraints on the use of federal funds for stem cells was not shared by all of his party — but not all Democrats favour overturning the veto. There is thus more reason than ever for the research community to furnish members of Congress with opportunities to educate themselves on the issues — an effort that, although it may take time, could prove vital to progress. What's more, if Bush's veto was an attempt to establish stem-cell research as a wedge issue that would help at the polls, it failed fairly comprehensively, so that opposition might conceivably be open to negotiation.

What the Democrats should not be expected to do is provide significant increases in funding overall — something they have both the power and the temptation to do. The damage done to America and the rest of the world by unsustainable deficits is real, and any lack of zeal in facing this problem would be a mistake. In that context, this would be a good time for Congress to look again at Bush's plans for NASA to re-establish a human presence in deep space. The outgoing Republican Congress gave its Republican president too much benefit of the doubt on this undertaking. The new Congress must, at the very least, articulate more convincing reasons than have yet been heard for such a colossal expenditure. ■

Order for microbes

Burgeoning microbial gene data require coherent efforts to make them readily usable.

Microbes don't subscribe to the single life. They are coupled with complex ecosystems of diverse, mutually dependent species. This complexity and the vast numbers of microbes in the ocean, the soil, in our gut and almost everywhere else pose a challenge to those seeking to understand microbial ecology.

In the 1980s, surveying the microbial world by sequencing the collective ribosomal RNA opened up new avenues. For the first time it was possible to get a glimpse of the make-up of complex microbial

systems. It's a reasonable assumption that the more similar these sequences are, the more closely related the microbes are, and the more closely related their lifestyles must be — hence the pursuit of insights into what microbes might be doing in their environments.

But this assumption turned out to be fragile, as it emerged that microbes frequently shuffle around their genes both within and between species. A similarity in one gene does not necessarily correlate with the absence or presence of other genes in the genome.

Fortunately, the continuous decrease in sequencing costs allows today's microbiologists to sequence not only a single gene from each of the most abundant species in a microbial ecosystem, but also, at least in theory, all the genes present. These composite genomes, or 'metagenomes', provide a wealth of information that could only be dreamt of even a couple of years ago. With sequencing facilities



continuing to increase their capacities by applying new technologies, and funding agencies supplying the necessary resources, sequencing the ocean or the contents of the human gut has become relatively easy. But how to extract meaningful information from a metagenome, and to gain insight into both the individual species' impact on the microbial community and the impact of this community on the ecosystem?

We can hope to unravel the function of every gene when individual species can be cultivated and genetically manipulated in the laboratory, but this is impossible when dealing with a complex community containing hundreds or thousands of species. Functional assignment of genes needs to be performed, even when the only information available is a string of nucleotide bases.

There are numerous databases and websites, public and not-so-public, some adhering to an easily understandable framework of standards and regulation, and some not so transparent. Five years ago it was a big disappointment to compare one's chosen sequence with the GenBank database and not find a 'hit'. Today there is a feeling of sheer inadequacy in the face of vast quantities of sequence and annotation

information — and an acute need for a degree in bioinformatics.

Publication in most cases (including the *Nature* journals) requires the deposition of sequence data into the GenBank or EMBL databases. Much less effort is spent depositing unpublished data or updating information that is already published. In all probability, in the not too distant future, metagenomic studies will be done not only by the big sequencing centres, but by anybody with a reasonable research budget and university support. To make all the data more easily accessible, it would be desirable to have a collaborative effort of genome centres and funding agencies to build a universal microbial-sequence database, with a readily comprehensible framework for sequencing and annotation standards and regulations.

Microbiology has come a long way from investigating the easily cultured individual microbe from a rich microbial community and describing what is out there, and is now starting to get a grip on what they actually do. With the intrinsic difficulties of dealing with complex systems, it is good to see a field galvanized by new technologies and scientific daring. But more infrastructural order is required, to prevent the discipline getting ahead of itself. ■

Success and successor

Exit an outstanding science minister; enter a more political operator.

Earning the respect of the sector you represent isn't always a priority for a government minister, and can even lead your prime minister to suspect that you've 'gone native'. But when Tony Blair (to his credit) appointed David Sainsbury to be Britain's minister for science in 1998, he will have known that Sainsbury was native already, having a deep respect for science and having given it many millions of pounds of his personal philanthropy. Precisely because of his commitment to science, Sainsbury wanted that job and no other in government. That simple fact, combined with Sainsbury's quiet yet successfully determined support for science throughout his eight years, underlies the chorus of almost unmitigated praise from leading UK scientists following his resignation last week.

The disappointing aspects of this era mostly happened despite Sainsbury, rather than because of him. The 'spin' in government statistics — whether in budgetary announcements or in statistics of growth in undergraduate numbers — is a part of Blair's legacy for which his probable successor as prime minister, chancellor of the exchequer Gordon Brown, shares responsibility and would do well to discourage. As a government minister, Sainsbury cannot be wholly absolved from responsibility for something he nevertheless tried to resist: the relentless growth of academic bureaucracy, supposedly in aid of accountability. It no doubt runs counter not only to his sense of what researchers should be focusing on, but also to his experience, based in his family's supermarket empire, of how to run a business.

Sainsbury's strengths have included making the time to engage with scientists and to understand their goals. He has been influential internationally in his support for a politically untrammelled Euro-

pean Research Council, and nationally in the way he has opposed and undermined vicious animal-rights activists. His time has been marred by accusations of conflicted interests, not least through hefty donations to a Labour party currently under police investigation for its handling of such support. But few believe that his reputation for integrity will be undermined when the results of that investigation are made public.

Sainsbury was luckier than most of his predecessors in that both his prime minister and his chancellor have been vigorous champions of science. But Sainsbury has had to defend his corner all the same. Now, the Treasury is focusing on critically analysing the benefits emerging from an unprecedented ten-year programme of annual increases in science funding.

Just what those increases amount to, given the exaggerations of spin and the pressures from commitments such as Iraq, education and public health, and just what benefits are demanded by the Treasury, are key issues for future scrutiny. Negotiating such challenges will be the responsibility of Sainsbury's successor, Malcolm Wicks — a member of parliament rather than a life peer and, by both definition and character, more of a political operator. Encouragingly, Wicks, in his current ministerial responsibilities for energy (which he will leave behind), has shown himself to be a good listener, prepared to consider evidence, and prepared to engage with lobbyists — yet also to face them down in public.

The respect and experience that Sainsbury has gained will stand him in good stead as he tackles his next commitment for the government: developing a strategic review of UK science policy. In this role, and in his position in the House of Lords, where he can scrutinize government legislation, and as a philanthropist, he can still influence science, and has earned the right to do so. ■

"Sainsbury's strengths have included making the time to engage with scientists and to understand their goals."

RESEARCH HIGHLIGHTS

Prickly all over

Science **314**, 941–952; 960–962 (2006)

The genome of the purple sea urchin, *Strongylocentrotus purpuratus*, has been sequenced. This will hearten developmental and systems biologists for whom the sea urchin serves as a model organism.

Sea urchins, like humans, are members of the superphylum of deuterostomes and hence are more closely related to us than the other lab favourites fruitflies and nematodes. The transparent sea urchin embryos are also ideal for studying early development. The sequenced genome reveals that many genes thought to be unique to vertebrates crop up in the sea urchin too, including some important for human hearing and sight. The genome also boasts large families of genes involved in innate immunity. Around 12,000 of its more than 23,000 genes are active in the developing embryo.



J. ROTMAN/NATUREPL.COM

BIOMECHANICS

Jeepers creepers

Phys. Rev. Lett. **97**, 184302 (2006)

How do climbing plants grasp the poles up which they wind? And why do some twine around thin poles but not thick ones? Researchers going back to Charles Darwin have puzzled over these questions. Now Alain Goriely at the University of Arizona in Tucson and Sébastien Neukirch at the Pierre and Marie Curie University in Paris, France, show that a simple model can explain the plants' behaviour.

They treated a twining plant stem as an elastic rod that has an intrinsic tendency to curl. They examined how friction between the rod and a supporting pole holds up the growing helix, and showed that the angle at which the curling tip meets the pole's surface determines the maximum radius of pole around which the stem can wind.

CELL BIOLOGY

Spaced out

Nature Struct. Mol. Biol. doi:10.1038/nsmb1170 (2006)

A molecular machine that seems to act like a ruler in fact performs a dynamic balancing act, researchers report.

Biologists know that the enzyme ACF controls the spacing between DNA-packaging units called nucleosomes, in which DNA is wrapped around protein spools.

To find out more, Geeta Narlikar and her colleagues at the University of California, San Francisco, tagged nucleosomes with fluorescent dyes and tracked their

movements in real time. It seems that ACF continuously samples the stretches of DNA bordering each nucleosome, and pushes longer pieces of DNA through the nucleosome faster than shorter stretches. This establishes an equilibrium state between neighbouring nucleosomes that favours even spacing.

PLANETARY SCIENCE

Time heals

Earth Planet. Sci. Lett. **251**, 79–89 (2006)

Planets such as Mars and Venus (pictured below) with rigid, unmoving crusts can experience transient bursts of plate tectonics, according to new computer simulations.

Alexander Loddock of the University of Münster in Germany and his colleagues built a model to simulate mantle convection beneath a 'stagnant lid'. Under some conditions the lid, or crust, breaks up and its parts begin to move. Previous studies have suggested that once such plate tectonics

starts, it either continues or recurs in episodes so closely spaced that the crust remains unsettled. Loddock's group found instead that bouts of plate tectonics can be far enough apart for the crust to completely reform. This would fit some theories of Mars' and Venus' histories, the team suggests.

MICROBIOLOGY

Burrowing bacteria

Science **314**, 985–989 (2006)

Researchers in Japan have figured out how *Shigella* bacteria, which cause dysentery, burrow through cells. Their observations may help efforts to combat the microbe, which kills 1.1 million people each year.

Chihiro Sasakawa of the University of Tokyo and his team used fluorescence- and electron microscopy to photograph cells infected with *Shigella*. The bacteria's movement is hindered by the cell's dense network of microtubulin, a cytoskeletal protein.

But *Shigella* produces a protein-chomping enzyme called VirA, which acts on microtubulin. Sasakawa's team saw that VirA sliced up the microtubules, creating tunnels through which the bacteria could spread. Knocking out VirA production in *Shigella* prevented the deaths of infected mice.

CELL BIOLOGY

Battle of the bulge

J. Cell Biol. **175**, 477–490 (2006)

Researchers hoping to understand a part of the cell's skeleton known as the cortex have made headway by studying a process called 'blebbing'.

The cortex is a layer of cytoskeleton



NASA/JPL-CALTECH

wrapped beneath the cell membrane, which helps to maintain cell shape and aid cell movement. Blebbing occurs when the membrane detaches from the cortex and inflates, forming a bulge on the cell surface.

Guillaume Charras and his colleagues at Harvard Medical School in Boston, Massachusetts, studied how the cortex reassembles beneath the bulging membrane. They show that it happens in steps, with actin and actin-bundling proteins following the anchor protein ezrin. Contractile proteins then pull the layer down, all within 30 seconds.

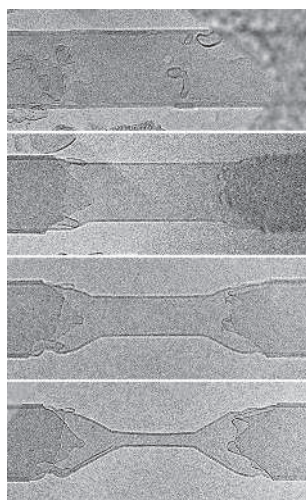
NANOTECHNOLOGY

Shrink to fit

Nano Lett. doi:10.1021/nl061671j (2006)

Carbon nanotubes shrunk to order could be the next fashion in nanotube devices, thanks to a process developed by Alex Zettl at the University of California, Berkeley, and his colleagues.

They blasted a nanotube with an electron beam, knocking some atoms from its structure. At the same time, a current was passed through the damaged tube. This made it reform, with few defects, at a smaller diameter (pictured right). The appeal for device designers is that a tube's electrical properties are intimately linked to its diameter: as the tube shrinks, resistance increases.



with the number that died in the surrounding tissue. For antiprotons, this ratio was some 3.75 times greater than for protons.

But practical therapeutic uses are far off: making antiprotons currently requires a circular accelerator with a diameter of at least 100 metres.

NEUROSCIENCE

Memory gene

Neuron 52, 437–444; 445–459; 461–474; 475–484 (2006)

Four papers published in *Neuron* help to demystify the mechanism of a gene implicated in the consolidation of memories.

The gene, known as *Arc/Arg3.1*, is expressed in the brain during learning. It has long been used as a marker of neuronal activity, even though its physiological role was not clear. The new research shows that mice with the *Arc/Arg3.1* gene knocked out fail to form long-lasting memories. Studies *in vitro* suggest that the gene controls the appearance and disappearance of receptors for the neurotransmitter AMPA on neuronal surfaces. Such receptor trafficking is known to modify the strength of connections between neurons, which is fundamental to learning and memory.

ASTRONOMY

Recipe for X-rays

Mon. Not. R. Astron. Soc. 372, 1531–1539 (2006)

On the outskirts of the spiral galaxy M99 is an object that, in some X-ray frequencies, outshines all the rest of the galaxy put together. It has been suggested that such 'ultraluminous X-ray sources' are powered by black holes larger than can be formed in the death throes of any normal star.

Roberto Soria of the Harvard-Smithsonian Center for Astrophysics in Cambridge, Massachusetts, and Diane Sonya Wong of the University of California, Berkeley, note that radio images of M99 show a cloud of fast-moving hydrogen impinging on the galactic disc near the X-ray source. They suggest that this influx of hydrogen could have compressed material at the edge of the disc into an anomalously large star, which later evolved into a black hole of a size suited to such X-ray brilliance.

JOURNAL CLUB

Michael Sanderson
University of Arizona, Tucson

A biologist turns his attention to evolution's neglected radiations.

Evolution's spectacular adaptive radiations get a lot of press: Darwin's finches and the Hawaiian silversword plants being textbook examples. These organisms, in adapting to environmental pressures, underwent both rapid speciation and radical morphological change.

Such episodes give rise to easily observable diversity and have stimulated extensive study. But how about those hyperdiverse clades in the tree of life in which many species have little morphological difference between them?

I first pondered this problem when musing about my thesis on the flowering-plant taxon *Astragalus*. I was cursed with perhaps 2,500 species, many remarkably similar. Their small differences were typically of uncertain adaptive significance.

Alas, I have counted barely ten papers since then that have addressed such radiations, which end up being labelled as 'non-adaptive'. I hope the most recent will shake things up a bit.

It analyses the speciation rate of North American *Plethodon*, a clade of salamanders most diverse in the woodlands of the Appalachian mountains (K. H. Kozak *et al. Proc. R. Soc. B.* 273, 539–546; 2006). This group has an evolutionary history that runs back 28 million years and has spun off about 46 species, many of which are only diagnosable by molecular markers.

Remarkably, the rate of speciation in the group's early days matched or exceeded rates seen in the textbook adaptive radiations. This suggests that we have a lot to learn about the evolutionary phenomena driving such radiations.

The authors make some interesting suggestions about the role of geography, ecology and adaptation in the salamanders' evolution. For example, the lineages may have evolved by tracking the ebb and flow of favourable habitats.

SPECIAL REPORT

The politics of breathing

Both sides in a US pollution dispute claim that science is on their side. **Emma Marris** explains how environmental laws have forced them into this position.

Last month, the US Environmental Protection Agency (EPA) made a ruling on air pollution that contradicts the recommendations of its own staff scientists and a panel of external experts. Many groups, from the National Resources Defense Council to the American Lung Association, are protesting loudly. Both the protesters and the EPA claim that science supports their case, but neither side is really right.

All parties insist that concern for human health is the only factor influencing their decision. But in fact they include different external factors, such as cost, in reaching different figures for the acceptable level of pollution.

At issue here are small particles found in the air, often the product of combustion in a power plant or car. Formally called 'particulate matter', they are more generally known as soot. When breathed in, the particles inflame the lung tissue, and can lead to obstruction of the airways, and heart and lung disease.

Periodically, the EPA comes up with acceptable annual and 24-hour exposure levels for particles smaller than 2.5 micrometres across (fine particles) and for particles smaller than 10 micrometres (coarse particles). Many countries and the European Union (EU) regulate coarse but not fine particles (see 'Dirty cities'). The EU has set targets for coarse particles that — assuming that fine particles make up about half to three-quarters of the coarse ones — mean that fine particles would be regulated at around 10–14 micrograms per cubic metre annually by 2010.

In the United States, the annual standard for fine particles was set for the first time in 1997 at 15 micrograms per cubic metre. This year, when the EPA was set to reconsider the limit, many wanted it lowered. But when the rules were announced on 17 October, it wasn't.

Public-health experts consider this standard to be the most important — the smaller the particle, the deeper it gets into the lungs, and long-term exposure accounts for more deaths than short-term peaks. At least two large, well-respected epidemiological studies have documented the phenomenon in

various cities and are cited extensively by both sides in making the case for different levels of regulation¹.

Arden Pope, an environmental epidemiologist at Brigham Young University in Utah, was involved in both studies. He says there is no evidence that it is safe to breathe any level of particulate. The relationship between particulate concentration and illness or death is more or less linear: the worse the air, the worse your health. The studies suggest that for every additional microgram of fine particulate in the air, between 0.6% and 1.6% more people will die every year — that's a few tens of thousands of deaths a year in the United States.

Science on its own does not provide an obvious standard — pick any number, and a level set below it will produce fewer deaths — but a number must nonetheless be chosen. "There is no threshold at which some sensitive group is not harmed," says Pope. "The judgement then becomes: is there some level at which we quit trying to go lower? Then it becomes a policy judgement."

Finding a level

That judgement is made by EPA administrator Stephen Johnson. As part of the process, agency scientists pull together the available science and put some values on the table for discussion. An external panel of experts reviews this and makes recommendations: in this instance, to lower the annual standard to 14 micrograms per cubic metre. But in the end, Johnson maintained the level at 15 micrograms.

According to the final ruling, Johnson gave more weight to uncertainties in the data at lower exposure levels, and based the long-term standards on long-term studies, instead of also including some related data from short-term studies. Interested parties have until 17 December to file a complaint, or a lawsuit.

The panel of external experts was livid. "We took our tests very seriously and came up with a recommendation that the administrator didn't take," says Rogene Henderson, head of the panel and an air-pollution expert at the Lovelace Respiratory Research Institute in Albuquerque, New Mexico. "My concern is

"The judgement becomes: is there some level at which we quit trying to go lower?"



that the administrator chose not to take our advice, but other people's."

Some critics charge that in not lowering the standards, Johnson went along with the administration of President George W. Bush, who has close ties to powerful business interests in the oil and gas industry. The EPA estimates that it would cost an additional \$2.5 billion to

Dirty cities

Cities worldwide vary dramatically in how much particulate matter, or soot, their air contains. New annual US standards for particles less than 2.5 micrometres across, set at 15 micrograms per cubic metre by the Environmental Protection Agency, have recently come under fire.

City	Particulate level ($\mu\text{g m}^{-3}$)*
Beijing	71–99
Mumbai	40–55
Mexico City	27–38
Athens	26–36
Los Angeles	18
London	6–9

*Values for non-US cities are estimated as a percentage of the value for particulate matter of 10 micrometres or less.



US SCIENCE MEETS NEW PAYMASTERS

Democratic-led Congress could shake up funding for science agencies.

www.nature.com/news



Wehrum also points to a Supreme Court case in 2000, in which the court supported the idea that air-quality standards should be set “at the level that is ‘requisite’ — that is, not higher or lower than necessary”. In that case, the EPA was being sued by a group of industries who argued that the annual fine-particulate level was set too low. This year, any potential lawsuit is likely to come from the health and environmental groups, saying it is set too high. The particulate standard they are suing over is, of course, the same.

Political decision

Whereas the EPA's external experts wanted the level to be lowered to 14 micrograms per cubic metre, the American Lung Association plumped for a safe level of 12. But the World Health Organization (WHO) recommends setting the limit even lower, at 10 micrograms. The association isn't saying it is willing to accept 50,000 more deaths each year than the WHO; rather, it supports the level of 12 micrograms on political grounds, as the EPA was never willing to consider 10 micrograms per cubic metre.

Deborah Shprentz, a consultant to the association, says various analyses support a level of 12 micrograms per cubic metre, and fumes that “the EPA did not provide a robust scientific justification for leaving the standard at the current level. It seems like a betrayal of trust”.

The New York Times gave its 14 October editorial on this matter the title “Science Ignored, Again”, but this might more properly be “Scientists Ignored, Again”, as it was scientists' policy wishes, rather than scientific facts, that were disregarded. Any decision about what standard to set that considered just health would slide right down the linear relationship to zero.

Sixteen years ago, in a National Academies report³ on environmental decision-making, Boston University political-science professor Shep Melnick summed up what he called a uniquely American phenomenon: “The widespread hostility to the use of benefit–cost and risk assessment analysis,” he wrote, “is based on an absolutist health-only position that virtually no one is willing to embrace in the real world. To put it more bluntly, almost no one really believes what many informed people emphatically maintain in public.” ■

1. Pope, C. A. & Dockery, D. W. *J. Air Waste Mgmt Assoc.* **56**, 709–742 (2006).
2. Schoenbrod, D. *Regulation* **29** (3), 36–42 (2006).
3. Hammond, P. B. & Coppock, R. (eds) *Valuing Health Risks, Costs, and Benefits for Environmental Decision Making* (National Academies Press, Washington DC, 1990).

Governments struggle with setting pollution limits when even small amounts cause health problems.

clean the air to 14 micrograms per cubic metre, instead of 15, by 2020. But the same analysis also suggests that lowering the standard to 14 micrograms per cubic metre would save the economy an additional \$10 billion in health and visibility-related costs.

Many also charge that Johnson's decision violated the Clean Air Act, which says the administrator must set levels “requisite to protect the public health with an adequate margin of safety”. This vague phrase has been interpreted by Congress and the courts to mean that the EPA must protect the public's health without considering other factors, such as cost. The law does not say how many deaths are acceptable. Many pollutants other than soot, including some carcinogens, have a linear relationship between dose and health, so considering health alone would seem to demand that the agency set for all pollutants the politically and practically impossible standard of zero.

David Schoenbrod, professor of environmental law at New York Law School, argues that Congress knew about this problem in the wording of the act when it was passed in 1970. In a recent article² he quotes the act's

sponsor, Senator Edmund Muskie, as saying in 1977: “Our public health scientists and doctors have told us that there is no threshold, that any air pollution is harmful. The Clean Air Act is based on the assumption, although we knew at the time it was inaccurate, that there is a threshold.”

Congress passed the act anyway, Schoenbrod contends, so that it could make perfect decisions about banishing harm and let the EPA worry about the inevitable compromises. “The EPA, as a practical matter, must decide exactly how much health risk to tolerate,” he says. “The EPA weighs health risks against cost.”

Of course, the EPA cannot legally admit to doing this. Bill Wehrum, the agency's acting assistant administrator for air and radiation, says the new regulations were based only on science. But in that case, why not bow to the wishes of the staff scientists and lower the regulatory level? “We give absolute deference to what the science says, but reasonable minds can differ,” Wehrum says.

“The EPA did not provide a robust scientific justification for leaving the standard at the current level.”



P. WILLIAMS/WHO

Margaret Chan: China's winning candidate for WHO director-general has been criticized for her handling of Hong Kong's SARS outbreak.

WHO boss faces test of independence

China has covered up outbreaks of SARS and avian influenza, and baulked at sharing samples and data with the international community. But last week, China's candidate for director-general of the World Health Organization (WHO), Margaret Chan, won a landslide victory in the election to replace South Korea's Lee Jong-wook, who died in May.

Chan's election might herald greater Chinese openness and involvement in global health. Many are watching to see if, on taking office in January, she is as independent of her sponsors as she has promised. She has made women and Africa her priorities.

The WHO faces a crisis of identity and purpose. The organization used to be the only global-health show in town, but the past decade has seen a series of new players emerge. The Global Fund to Fight AIDS, Tuberculosis and Malaria has a bigger budget than the WHO's US\$1.7 billion, and the Bill & Melinda Gates Foundation spent almost \$1 billion last year supporting global health. And nearly 100 public-private health partnerships have sprung up, for example to develop drugs and vaccines for malaria and leishmaniasis.

In fact, the presence of more organizations spending more money gives the WHO a chance to reinvent itself. For all its bureaucratic

rigidity, it remains the only agency plugged into health ministries worldwide, with staff in almost every country, and the capacity to bring together the world's top scientists to tackle a particular problem. In principle, it is well-placed to help integrate the newer initiatives.

In her manifesto, Chan endorsed such a shift in the WHO's focus. She is also apparently open to change: she recently led a review panel that concluded that the Tropical Disease Research programme sponsored by the WHO, the United Nations and the World Bank needs to be rethought. Yet Chan's victory surprised even some of her colleagues, who say that her track record, latterly with the WHO's programmes on communicable diseases and pandemic influenza, is more as a diplomat than a leader.

"She is untested on the technical and political side," says Richard Horton, editor of *The Lancet*, which backed another candidate. "One of her first tests will be who she brings in: she needs a team that fills the gaps in her own experience."

On the eve of the election, China hosted a summit of African heads of state, and announced more than US\$10 billion in aid for trade, health and agricultural development. Besides China, the United States backed her bid. Her 12 rivals included Bernard Kouchner,

former French health minister and founder of Médecins Sans Frontières.

Chan campaigned on her "outstanding leadership" as director of health in Hong Kong during the H5N1 avian flu outbreaks in 1997 and the 2003 SARS outbreak. But in 2004, a Hong Kong government inquiry sanctioned Chan for "unsatisfactory performance" in her handling of the SARS outbreak. And Hong Kong's cull of all chickens in late 1997 to stop the spread of H5N1 happened seven months after the first human H5N1 cases were reported.

Most observers are giving Chan the benefit of the doubt. "She has travelled the rocky road of H5N1 and SARS and is well equipped to head the WHO in the many challenges and opportunities that lie ahead," says Ken Shortridge, a veteran of infectious-disease research in China, now retired in New Zealand.

For Ellen t'Hoen, who leads Médecins Sans Frontières' campaign for access to essential medicines, Chan's priority must be to put health above national and business interests. For months, under US pressure, the WHO has sat on an advisory document on using trade rules to negotiate cheaper drugs. Chan would demonstrate her independence by publishing the document, says t'Hoen.

Declan Butler

Q&A: BART GORDON

Leadership is also shifting in Washington DC, where the 7 November elections ushered in a complete changeover. Starting in January, Democrats will replace Republicans as the majority party in both houses of Congress. As such they will lead all committees, the working groups that regulate the details of laws in progress. *Nature's* **Emma Marris** spoke to one of the new power brokers in Congress: Representative Bart Gordon of Tennessee, who as ranking Democrat is in line to become chair of the House Committee on Science. She asked him how things will be different for science issues in Washington DC.

What is the main difference between you and your predecessor as chair of the science committee, Sherwood Boehlert?

Sherry's leaders in the House unfortunately limited what he could do, so Sherry couldn't be Sherry, if you know what I mean. I don't think that we are going to have that kind of heavy-handedness from the Democratic leadership.

What's going to happen with the president's American Competitiveness Initiative, to keep the country competitive in science and technology?

We will push legislation forward on that. I am very excited about it. I don't want this to be a Democratic or a Republican issue. I was asked by [incoming speaker of the House of Representatives] Nancy Pelosi to work on a competitiveness agenda, and this is a high priority for her.

One proposal that you have put forward is the creation of an Advanced Research Projects Agency for Energy, to work on technology that will reduce oil imports. Where is that going?

I don't think that the administration has an interest in it. They want to do things the way they want to do it and don't really want anybody else making recommendations. Energy independence is very important, I think it makes sense, and I am going to push very hard, in what I hope will be a bipartisan way, to get this accomplished.

There are noises from the administration that they are to release a sweeping energy plan. Are sweeping changes the way to go?

I hope that what you will see from the science committee is a lot of narrow-issue, bipartisan, good-government initiatives. My thought is that we could sit around here and talk round and round about energy independence for

two, or three, or four years, and not get anything done. I don't want to hold up any good ideas we might have to put into a major bill if we can get them through on their own. Let's do smaller steps that we can get done.

Give me an example.

We are going to start with the education initiatives. There is a consensus that the United States has got to raise its capabilities in math and science. If we are going to have the jobs of the future, we need a higher-skilled workforce. The real problem is that approximately two-thirds of the math teachers in this country have neither a major nor a certification in that area. Compared with other countries, our students do worse the longer they are in school. The National Academies report lays out programmes to change this, and I'd like to see them put in place.

Does NASA have its priorities right? Do you feel that a realistic number has been put on the cost of sending humans to Mars?

I would like to see NASA do all that it is proposing and more. But we need to do a better job of oversight. I want to see if all of the numbers add up, and frankly I don't think they will. If they don't, we will have to take a hard look at priorities. I don't want to pass problems on to others. What we have seen with NASA is that prior administrators just keep on passing on problems. Someone needs to take oversight.

You've talked about censorship of scientists at agencies and politicization of science. Do you plan any hearings on that issue?

We hope that we can have some oversight hearings that are going to find out what was really going on. My goal is not to embarrass the administration but to shed some light on this problem so that people will be embarrassed to do it again. ■

See Editorial, page 243.

ON THE RECORD

"We need our scientists today to be as celebrated and famous as our sportsmen and women, our actors, our business entrepreneurs."

UK prime minister Tony Blair, speaking in Oxford, 3 November.



E. AGOSTINI/V. ARMAND/AFP/GETTY

NUMBER CRUNCH

1 is the number of scientists in *Time International's* 60 years of heroes (Andrei Sakharov, right)

9 is the number of musicians (John, Paul, George and Ringo, Maria Callas, John Lydon (left), Fela Anikulapo-Kuti, Bono, Bob Geldof)

SCORECARD



Hope
HOPE (Help Ohio Public Education) got four pro-evolution candidates elected to Ohio's board of education. And Dick DeVos, a Republican gubernatorial candidate in Michigan who wanted 'intelligent design' taught in school, was trounced.



Flares
It sounds '70s retro, but astronomers at a workshop on Cool Stars heard of what might be the biggest magnetic flare ever. An outburst in the II Pegasi system spotted by NASA's Swift satellite was 100 million times more energetic than a typical solar flare.



Space-shuttle hootenannies
NASA is keeping the shuttle Discovery grounded over the New Year, because some of the shuttle's software cannot reset itself from 365 to 1. Rebooting is not a preferred option when in orbit.

Sources: The Times, New Scientist

UK civil servants accused of warping science

"Politicians twist science to suit policy" would be an unsurprising headline in the United States, given the rocky relationship between the Bush administration and the US science community. But the same message raised eyebrows on the other side of the Atlantic last week, when a UK parliament report suggested that politicians there also pick and choose scientific results that best suit their policies.

Britain's three successive Labour governments have made much of their desire to integrate research into policy-making; phrases such as "evidence-based policy" have become buzzwords for ministers. Science minister David Sainsbury, who left his position last week after eight years, also won the respect of researchers (see page 244). But the latest report from the House of Commons Select Committee on Science and Technology tarnishes that happy image.

The cross-party group of 11 MPs takes ministers to task for labelling policies "evidence-based" when no relevant research exists, and criticizes the civil service for its poor interpretation of research results. Perhaps most worrying, concludes Phil Willis, the Liberal Democrat member who chairs the committee, is the fact that government-commissioned studies regularly go unpublished when they conflict with a department's policy.

The report's most dramatic examples came from Tim Hope, a criminologist at the

University of Keele, whom the Home Office commissioned to evaluate an initiative on reducing burglary. Hope told the committee that the Home Office ignored a result showing an increase in crime rates in one area, and focused solely on a second result that showed a drop in offences. The Home Office did not return *Nature's* request for comment.

Willis says that this and other examples — such as a 2005 anti-obesity drive developed in the absence of any evidence that it would work — show that the government lacks the academic ideal that all results must be aired, regardless of whether they fall as desired. "It's a level of scientific incompetence," he says. "There is not the culture of using scientific evidence and research in the way the scientific community would understand it."

The report recommends that the government should ensure that every department has its own chief science adviser, and should also establish a government scientific service charged with bringing scientists into government and securing proper career paths for them.

That is a laudable but old idea, notes Peter Cotgreave, director of the London-based lobby group the Campaign for Science and Engineering; a similar recommendation was made more than seven years ago, yet remains to be implemented. The government is likely to respond to the current proposals in the next few months.

"There is not the culture of using scientific evidence in the way the scientific community would understand it."



J. FOX/ALAMY

The report also suggests that peer review by outside bodies, such as learned societies, could assess the degree to which the government's policies are based on evidence. The government may, however, balk at the thought of having an outside body audit its performance. Some science-policy experts are also sceptical about the idea: "This would replace judgements currently being made by officials subject to democratic accountability with judgements made by those outside the process," points out Roger Pielke Jr of the University of Colorado, Boulder.

Whatever the difficulties in Britain, political interference in science policy is far greater in the United States, says Willis. For instance, it is widely thought that David King, the UK chief

Gunmen seize academics at Baghdad ministry

As *Nature* went to press, agencies were struggling to confirm details of what may be one of the worst mass kidnappings since the Iraq conflict began in March 2003. At around 9:30 a.m. local time on 14 November, gunmen are reported to have abducted up to 150 academics, staff and visitors from an office of the higher-education ministry in the Karrada area of Baghdad.

Little information on the institute targeted by the kidnappers was

available at the time of writing, although sources in Baghdad said that the centre involved is a branch of the ministry that helps students and professors obtain placements in overseas universities. Some of the staff would have had a scientific background. The identity of the kidnappers isn't known, but those taken include both Shia and Sunni Muslims. According to some reports, the higher-education minister, Abed Theyab, immediately ordered

the closure of all universities until security is improved.

The timing may have been chosen to coincide with the Pugwash Conference on Science and World Affairs, held 11–15 November in Cairo, Egypt, according to Abbas Al-Hussaini, a civil engineer at the University of Westminster, UK. Al-Hussaini is also general secretary of the Iraqi Higher Education Organising Committee in London, which was set up in January 2004

to help reconstruct Iraq's research and higher-education system.

The Pugwash organization, which campaigns against armed conflict, depends heavily on academics for its work in the Middle East, as well as other regions. Al-Hussaini believes that the kidnappings are not related to religious conflict between Sunnis and Shias, but are part of a plan to destabilize Iraq coordinated by former intelligence officials loyal to Saddam Hussein.



Selective evidence: data on crime initiatives were ignored by the UK Home Office.

science adviser, would be able to speak out if government science policy diverged sharply from the evidence, whereas few would expect such independence of US presidential science adviser John Marburger. But during King's six years in office, government policy and scientific thinking on high-profile issues such as climate change have not diverged substantially. An interesting test of his independence, and the government's commitment to evidence-based policy, will occur when they do. ■

Jim Giles

German stem-cell law under fire

Germany's strict laws governing human embryonic stem (ES) cells are no longer appropriate and need to be relaxed, says the country's main funding agency, the DFG.

Backing its arguments with an 80-page report released on 10 November, the DFG argues that its previous support for a cautious approach is no longer valid.

When the DFG speaks, politicians listen, and a parliamentary debate is now likely. But federal research minister Annette Schavan, a Christian Democrat, swiftly rejected any fundamental change to the rules, which forbid German researchers from working on human ES cell lines created after January 2002. The penalty for doing so, either in or outside Germany, is up to three years in prison.

The DFG calls for three changes to the law. First, that the cut-off date be removed to give researchers access to the newer, better stem-cell lines used in other countries. Second, that human ES cell lines be allowed to be imported for clinical as well as research purposes. And third, that the threat of punishment for German researchers working abroad be lifted.

Despite Schavan's rejection of the recommendations, there are splits on the issue for the first time in both her party and its sister party, the Christian Socialists, with some members calling for the cut-off date to be abolished. The Social Democrat coalition partner and the opposition Free Democrats are also split.

In previous statements in 1999 and

2001, the DFG called for continuing public debate about the possible benefits and limitations of stem-cell research, and for further research into the potential of adult stem cells to provide an alternative source of cells capable of generating different types of tissue. Its current report acknowledges that the past five years of international research has not only cast doubt on the potential of

adult stem cells, but has also made clinical applications of human ES cells foreseeable. German researchers are being left behind, the report says, and isolated further by a reluctance abroad to include German researchers on international stem-cell committees for fear they may be prosecuted at home.

The outcome of the debate is uncertain, but politicians are broadly supportive of

decriminalizing research on human ES cells by German scientists in countries where it is allowed.

Responding to the DFG's report, Schavan also promised to launch "in the near future" a research programme for alternatives to human ES cells. But this just annoyed researchers more. "This top-down attempt to provide alternatives has been the ministry's line since the beginning, and has been shown not to work," says Oliver Brüstle, head of the Institute for Reconstructive Neurobiology at the University of Bonn, one of the DFG report's 12 authors. "It will be extremely dangerous to ignore international developments," he warns. ■

Alison Abbott



Annette Schavan is against changing Germany's rules.



INDIA'S MONSOONS CAUSE MORE FLOODING

Rains hitting in the wrong place and time bring disaster.

www.nature.com/news

AMIT DAVE/REUTERS

T. SCHWARZ/REUTERS

MAHMOUD RAOUF MAHMOUD/REUTERS



Security has been tightened at Iraq's higher-education ministry.

If the details are confirmed, the abduction will be the biggest single event in a steady campaign of

attacks and assassinations against Iraqi academics during the bloody aftermath of Saddam Hussein's fall. More than 200 are thought to have been killed, and hundreds more have fled the country (see *Nature* 441, 1036-1037; 2006).

Lack of investigation and prosecutions means that little is known about who carries out such attacks, and the motives are thought to vary. Some victims have certainly been targeted in revenge for past political allegiances, but many believe that there is also an organized campaign to eliminate

intellectuals, as part of an attempt to make the country ungovernable.

"Terrorist forces are out to scare the scientific community," Al-Hussaini told *Nature* earlier this year. He believes that academics are targeted because they enjoy "much greater prestige and status than in the West, and could transform Iraq into a modern society".

Scientific and human-rights organizations, including Scholars at Risk, the International Council for Science and the American Association for the Advancement of Science, have called for

assassinations and attacks to be better investigated, for security to be boosted at universities, and for those most at risk to be given asylum at universities abroad.

But the killings continue. Among the most recent murders was that of Essam al-Rawi, a geologist and president of the University Professors' Union, who was shot dead on 30 October. A few days later, gunmen killed Jassim al-Asadi, a dean of the University of Baghdad, along with his wife and son. ■

Jim Giles and Declan Butler

Neanderthal genome sees first light

Two research teams publish the first detailed glimpse of a Neanderthal genome this week, starting the race to unravel the genetic structure of modern man's closest relative. Articles in *Nature*¹ and *Science*² offer complementary yet contrasting views of nuclear DNA extracted from the same bone of a Neanderthal male who lived 38,000 years ago, found in a Croatian cave.

Both groups, which collaborated in part on their projects, say they need to sequence far more of the Neanderthal's DNA to answer key questions. But for now, the studies find roughly similar timescales for when Neanderthals and modern humans, *Homo sapiens*, diverged from each other in the course of evolution.

Given recent advances in DNA sequencing, a full Neanderthal genome could be sequenced within two years, the *Nature* authors say. "This is the golden age of ancient DNA work," says first author Richard Green, a postdoctoral researcher at the Max Planck Institute for Evolutionary Anthropology in Leipzig, Germany.

Neanderthals and *H. sapiens* roamed the same regions of Europe and western Asia until the Neanderthals went extinct 30,000 years ago. Both research groups now have genetic estimates for when the two species shared a most recent common ancestor.

In the *Science* article, team leader Edward Rubin places that event at roughly 706,000 years ago. In *Nature*, the Max Planck group — headed by palaeogeneticist Svante Pääbo — write that human and Neanderthal sequences diverged roughly 516,000 years ago. Within the large margins of error typical of such analyses, these conclusions could turn out to be the same. The *Nature* paper also takes a stab at estimating the size of the ancestral population that gave rise to humans and Neanderthals: roughly 3,000 individuals.

The gene studies may also shed light on a long-contested idea in palaeoanthropology: that Neanderthals and *H. sapiens* may have interbred. Last week, another research team proposed that modern humans acquired a version of the *microcephalin* gene — which regulates brain size — through interbreeding with another *Homo* species that has since gone extinct, possibly the Neanderthals³. Fossils — such as a 30,000-year-old skull from a Romanian cave⁴ — have also been put forward as evidence of hybridization between the two species.

The latest data suggest that some interbreed-



Sequencing of DNA from ancient bone might result in a full Neanderthal genome within two years.

ing may have taken place on a limited scale, says Pääbo. "One possibility is gene flow from modern human to Neanderthal," he says. But the data could also be explained if the studied Neanderthal sample was contaminated with modern human DNA. "We have contamination under control," Pääbo says.

As for the other group, "we don't see any evidence of interbreeding," says Rubin, director of the Joint Genome Institute (JGI) in Walnut Creek, California. "But we could have missed it because we don't have enough Neanderthal sequence yet."

Rubin's group reports analysis of about 65,000 base pairs. The *Nature* article reports about 1 million. The genomes of modern humans and Neanderthals each have about 3 billion base pairs. "These studies show you can do large-scale sequencing of Neanderthal DNA," says Jeffrey Wall, a computational biologist at the University of Southern California in Los Angeles who was not involved in either study. "There are some unresolved issues," he says. "In the end, the questions will be answered to everyone's satisfaction."

Both groups agree that less than 0.5% of the two species' genomes are different. But the teams largely used different sequencing methods. The JGI team opted for a traditional bacterial, or Sanger, method to replicate Neanderthal DNA. The Max Planck team used a newer, and faster, sequencing system⁵.

Both groups are pumping new sequences

into publicly accessible databases. Nearly 5 million base pairs are already available. By spring, Pääbo expects to have about 1% of the Neanderthal genome completed.

Rubin is now developing methods to target specific genes in the regions of the genome that are different in Neanderthals and modern humans. This approach, he says, allows one "to reach in and fish out DNA for analysis for different traits".

One of the first fishing trips will be for a gene called *FOXP2*, which in *H. sapiens* is linked to speech and language. Many animals have the gene, but the human version has distinct differences from that of the chimpanzee⁶.

Such genetic excursions will need more Neanderthal individuals for a truly representative picture of the differences between the species. Palaeoanthropologists and geneticists are now teaming up to acquire more bone, from more caves.

Eventually, genetics and palaeoanthropology will intertwine, predicts James Noonan, a JGI postdoc and first author on the *Science* article. "How biological features are encoded in the genome," he says, "will be correlated with what is seen in specimens."

Rex Dalton

1. Green, R. E. et al. *Nature* **444**, 330–336 (2006).
2. Noonan, J. P. et al. *Science* **314**, 1113–1118 (2006).
3. Evans, P. D. et al. *Proc. Natl Acad. Sci. USA* doi:10.1073/pnas.0606966103 (2006).
4. Soficaru, A., Dobo, A. & Trinkaus, E. *Proc. Natl Acad. Sci. USA* doi:10.1073/pnas.0608443103 (2006).
5. Margulies, M. et al. *Nature* **437**, 376–380 (2005).
6. Enard, W. et al. *Nature* **418**, 869–872 (2002).

J. KRAUSE

Foreign-student enrolment bucks declining US trend

The total number of international undergraduate and graduate students enrolled at US universities held even this year. This marks the end of two consecutive years of decline, according to the most comprehensive survey of the 2005–06 school year.

Roughly 565,000 foreign students studied in the United States in that academic year, said the Institute for International Education, an organization based in New York that tracks enrolments in the United States of students from abroad. That number was down by a fraction of a per cent from the previous year.

Enrolments from South Korea were up by 10%, whereas those from China were flat and the number of students from India dropped by nearly 5%.

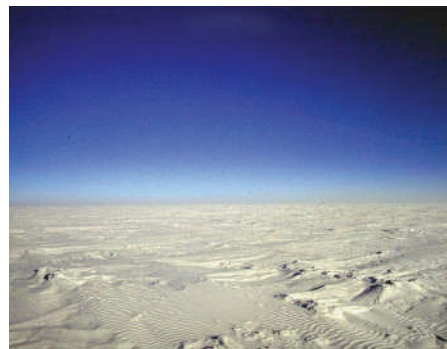
But first-time enrolments were up 8.3%. “The slide in enrolment has basically bottomed out,” says Peggy Blumenthal, an executive vice-president at the institute. “We’re turning the corner.”

China set to drill for Antarctica's oldest ice

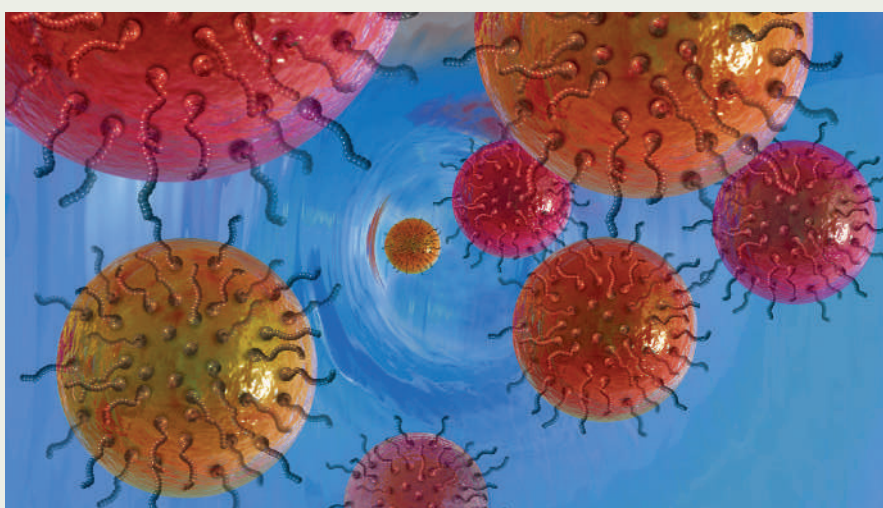
Things are looking good for China's search for the oldest ice in the Antarctic, climate researchers heard on 12 November at the Earth System Science Partnership meeting in Beijing. The group aims to drill deep into an ice cap called Dome Argus, 4,000 metres above sea level.

A partial analysis of the 110-metre-long core taken from the dome in a 2005 expedition (see *Nature* 433, 564; 2005) helps to confirm that snow has accumulated slowly in the area over the past 10,000 years at an average rate of 1.5 cm per year, says expedition member Xiao Cunde of the Chinese Academy of Sciences in Lanzhou.

That shows that the ice at the cap's bottom should be more than 1 million years old, or possibly even 1.5 million years old, said Cunde.



Frozen time: climatologists hope to extract 1-million-year-old ice from Antarctica's plateau.



CBEN/RICE UNIV.

'Nanorust' to provide safe drinking water for millions

A new technique to remove arsenic from drinking water could offer hope to the millions of people in India, Bangladesh and other developing nations who have no choice but to rely on wells that hold contaminated water. The method, developed by researchers in Texas, uses tiny rust particles, pictured above, that are smaller than viruses.

These 'nanorust' particles are stirred through the water and then dragged out using a simple hand-held magnet, report Vicki Colvin and her colleagues at Rice University in Houston (V. Colvin *et al. Science* 314, 964–967; 2006). The iron present in the rust (magnetite; Fe_3O_4) binds to the arsenic, reducing it to the safe levels stipulated by international drinking-water standards of the US Environmental Protection Agency. The team is now working on a cheap and simple manufacturing method using rust mixed with olive or coconut oil as a dispersal agent.

Indian space agency wants person on Moon by 2020

India's space agency, the ISRO, has announced plans to put a person in space by 2014 and to land an Indian on the Moon by 2020, four years ahead of China.

The decision reverses the ISRO's policy of using space technology for national development instead of for manned flights, which come at a huge extra expense. The agency will seek US\$2.2 billion in state funding — treble its current annual budget — for its manned flight programme, and many times more for the Moon landing.

The plans, unveiled to Prime Minister Manmohan Singh last month, were endorsed by a cross-section of the scientific community at a consultative meeting on 7 November in Bangalore. Formal government approval will be sought before the end of the year, says ISRO chairman Gopalan Madhavan Nair.

England's universities get more cash for chemistry

England's ailing chemistry departments, several of which have been closed in recent years, have been thrown a financial lifeline.

The Higher Education Funding Council for England (HEFCE), which distributes grants to universities for teaching and other costs, said on 8 November that it will inject £75 million (US\$143 million) over

the next three years into science subjects such as chemistry and physics, which are expensive to run. The University of Exeter and a handful of other institutions have closed chemistry departments recently, but the money is expected to prevent any more closures in the short term.

Science-policy experts are now calling on the HEFCE to recognize that teaching subjects such as chemistry is unusually expensive, and to change the criteria it uses to distribute money accordingly.

Mars orbiter performs anniversary stunt

A pre-celebration headache has struck NASA, which lost contact with the Mars Global Surveyor just days before the tenth anniversary of its 7 November launch.

The craft was having trouble moving its solar panels early in November, and a two-day silence followed. On 5 November, a signal was received showing that the orbiter had gone into a pre-programmed safe mode while awaiting instructions from Earth. As *Nature* went to press, scientists had still heard nothing since. They think the orbiter has adjusted its solar panels into a position that makes communication difficult.

If it does not receive any signals from Earth for seven days, the orbiter is programmed to transmit from its high-gain antenna. Should this signal not arrive, the spacecraft might have to stay missing in action.

BUSINESS

The chips are down

Geneticists' desire to track the roots of complex diseases has shaken up the market for gene chips.

Meredith Wadman reports on two firms jostling for position in a business potentially worth US\$500 million.

Wall Street analyst Elise Wang began her coverage of the Goliath of gene-chip makers last month with an unequivocal recommendation: sell.

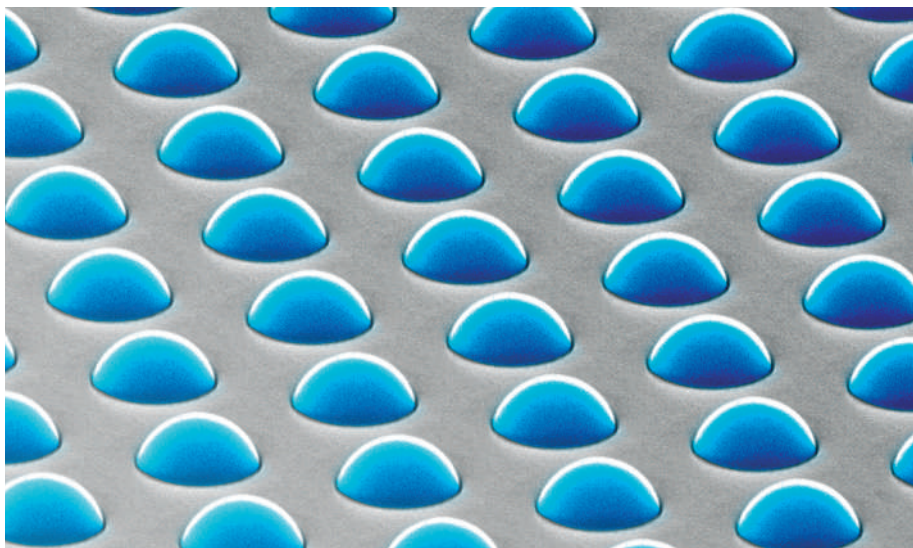
Her advice came despite the fact that shares in Affymetrix, the company based in Santa Clara, California, whose scientists invented the gene chip 17 years ago, are already trading down by almost 50% from where they were a year ago (see graph). "We believe Affymetrix may trade lower, especially as competition continues to intensify in the genotyping market," the Citigroup analyst advised her clients.

Opinions are a dime a dozen on Wall Street, of course, and analysts at UBS and Bear Stearns last month were neutral and positive, respectively, on Affymetrix. But whatever their take, few analysts dispute Wang's picture of a company locked in a major battle with an upstart rival: Illumina. The San Diego-based firm, founded in 1998, has half Affymetrix's number of employees but following the recent tripling of its share price, a larger market capitalization. The stakes in this battle aren't trivial. The two companies are fighting for the biggest growth opportunity in the biotechnology toolkit.

Gene chips, or microarrays, are wafers of silicon, glass or plastic that, when mounted with microscopic DNA probes, allow biologists to study when and how genes express themselves. The battleground today isn't over the use of gene chips to study gene expression — Illumina is a minor presence in that already-crowded field. There, Affymetrix generates more chip revenues than all of its competitors combined. But Affymetrix and Illumina are the only two players in the brand new market for high-density genotyping, valued conservatively at US\$225 million this year and estimated to be growing at between 35% and 45% annually.

Mapping success

In high-density genotyping experiments, researchers use microarrays to study DNA from thousands of biological samples from different people — some with a given disease and some without. Their aim is to identify in each sample hundreds of thousands of SNPs — single nucleotide polymorphisms. These are rare bases in the three-billion-base genome that vary between people. The hope is that the patterns of this variation are associated with common diseases, such as diabetes or asthma,



Chip wars: Illumina's bead-array technology provides competition to Affymetrix's 500K gene chip.

and with patient responses to drug treatments. The launch of scores of such experiments has been enabled by the HapMap Project — a major international collaboration that for the first time described common human patterns of sequence variation and whose first results were published in October 2005 (The International HapMap Consortium *Nature* **437**, 1299–1320; 2005).

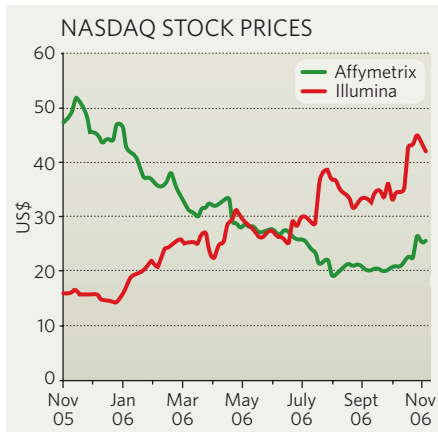
At that time, Affymetrix and Illumina were racing to capitalize on the market that the HapMap created. In the summer of 2005, Affymetrix had introduced its flagship high-density-genotyping product, the 500K: a set of two chips that it promised would reliably detect 500,000 SNPs in a biological sample. Using the

same technology employed to make semiconductor chips, but fixing each chip with DNA probes aimed at identifying 250,000 SNPs, the company introduced its star product as "the first practical and affordable solution for genome-wide association studies".

But Affymetrix's 500K soon stumbled, with a series of manufacturing and technical glitches — including a serious fault in a crucial software algorithm that did the genotyping. The result: scientists lost reams of data and the company was left busily replacing defective chips.

The problems "did introduce substantial delays" for investigators, says David Altshuler, a human geneticist at the Broad Institute in Cambridge, Massachusetts, whose scientists spent months helping the company rewrite the faulty algorithm. "A very large number of the chips they shipped last year were defective," adds John Hardy, a human geneticist at the US National Institute on Aging. "I suspect that they released it just too soon, to get out ahead of Illumina."

Sean George, vice-president of the academic business unit at Affymetrix, says that problems with the 500Ks are now behind it. "We shipped the 500K early because it was an exciting product and it has an unprecedented amount of genetic power," he says. "We've come out with a new algorithm and our customers have since scaled up into full production."



But before the release of the new algorithm in May, not much was going right. In January, just as Affymetrix's customers were grappling in frustration with the 500K's problems, Illumina released its competitor, the single-chip HumanHap 300K, which was soon followed by 550K- and 650K-SNP versions. Illumina's chips — with 'beads' that carry short DNA snippets, each matched to a particular SNP — had the advantage that knowledge obtained during the HapMap project was fed into their design.

Open market

In April, Affymetrix took the unusual step of withdrawing the guidance on its expected 2006 earnings that it had offered investors three months earlier — a sign that management was facing such difficulties that it was no longer able to forecast revenues. Affymetrix's problems coupled with Illumina's new releases spelt trouble for the more established company. Jay Flatley, Illumina's chief executive, claims it has won up to three-quarters of the market for chips used in high-density genotyping.

George disputes that assertion. And in the past few months, Affymetrix seems to be recovering somewhat. In the summer, it halved the price of its 500K chip set, to \$250, anticipating the introduction early next year of a new 500K product that runs on just one chip. Illumina's competing chip remains 30–50% more expensive. Affymetrix's move got the desired results: customer orders for the current 500K chip set surged by 90% in the third quarter, helping its stock climb from a 1 August low of \$17.50 to \$26 on 10 November.

In September, the company landed a deal to analyse more than 9,000 samples from the Framingham Heart Study, in search of SNPs associated with heart, lung and blood disorders. At the same time, Illumina has been landing contract after contract, most recently being chosen by Amgen to genotype 28,000 samples as part of the Women's Genome Health Study.

Affymetrix aims to follow the new 500K with a million-SNP chip by mid-2007. And in the spring, it is due to confront Illumina in a Delaware courtroom, where Affymetrix will try to persuade a jury that its arch-competitor is infringing five of its patents. A loss in the case would be a serious blow to Illumina: Flatley concedes that three-quarters of the company's revenues — \$124 million so far this year, compared to Affymetrix's \$251 million — could potentially be deemed as infringing one or more of the five patents at issue.

In the meantime, neither company will be standing still. They'll be looking to exploit another nascent gene-chip market in the diagnosis of specific diseases, an annual market some analysts think could one day top \$1 billion. ■

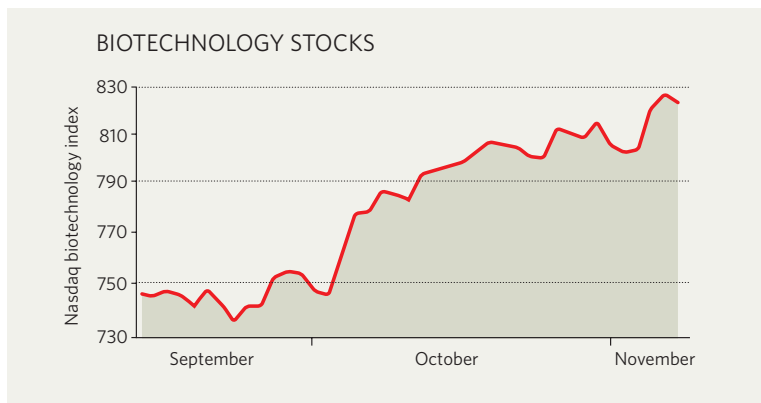
IN BRIEF

SHANGHAI BOUND Novartis says it will build a \$100-million research centre in Shanghai, specializing in infectious causes of cancer. The lab will start operations next May in temporary premises and then build a larger facility, which, the Swiss drug company says, will eventually house 400 scientists. The centre will be part of a network of eight strategic labs that Novartis operates around the world, and represents one of the most determined efforts yet by pharmaceutical companies to establish a research foothold in China (see *Nature* 440, 990–991; 2006).

AIRBUS DIP The parent company of Europe-based aircraft manufacturer Airbus reported sharp losses in the third quarter, caused in part by continuing delivery problems with the A380 superjumbo (see *Nature* 443, 385; 2006). Aerospace group EADS reported an operating loss of €239 million (US\$308 million), against profits of €559 million in the same period last year. Airbus suffered another serious blow on 7 November, when FedEx cancelled an order for ten freight-carrying A380s and said it would buy from Boeing instead.

TRANSLATION PUSH The University of Tokyo has announced that it will launch a Translational Systems Disease Biology initiative, which will try to use proteomic and epigenomic approaches to seek disease cures. It aims to work across departments and with outside companies to develop new drugs within ten years. The initiative is seen as part of a push by the medical school at Japan's leading university to break with its 'ivory-tower' image and forge closer links with patients and with business.

MARKET WATCH



This week Wood Mackenzie, an Edinburgh-based research and consulting firm, reviews recent trends in biotechnology stocks. Boosted by a better investment sentiment and a clutch of eye-catching acquisitions, the Nasdaq biotechnology index is up 11.5% over the past eight weeks — but is still less than 4% up from where it started the year.

Celgene of New Jersey was a prime mover in the biotech index, with its shares up almost a third after it reported strong sales of the anticancer drug Revlimid, as well as news that the company is to be listed on Standard & Poor's 500 index.

On 2 October, Californian drug maker Gilead agreed to buy Colorado cardiovascular company Myogen for \$2.5 billion, impressed by strong clinical results for Ambrisentan, a treatment for pulmonary arterial hypertension. Myogen's shares jumped 47% on the news. And later that month, Stiefel

Laboratories of Florida agreed to pay \$640 million for Connetics, a Californian dermatology company, whose share value rose 46% on the news.

Then, in perhaps the most prominent of the purchases, Merck said it would pay \$1.2 billion for Californian RNA interference company Sirna Therapeutics. Sirna is attempting to develop an RNAi drug to treat age-related macular degeneration. Sirna's shares almost doubled on the announcement.

And on 6 November, Illinois-based Abbott Laboratories bought Kos Pharmaceuticals of New Jersey for \$3.7 billion, with an eye to its cholesterol-lowering drug Niaspan. Kos's shares were up by more than half on the news. Like the other deals, this served to remind the markets that drug companies are ready and willing to pay a healthy premium to get their hands on novel biotech products. ■

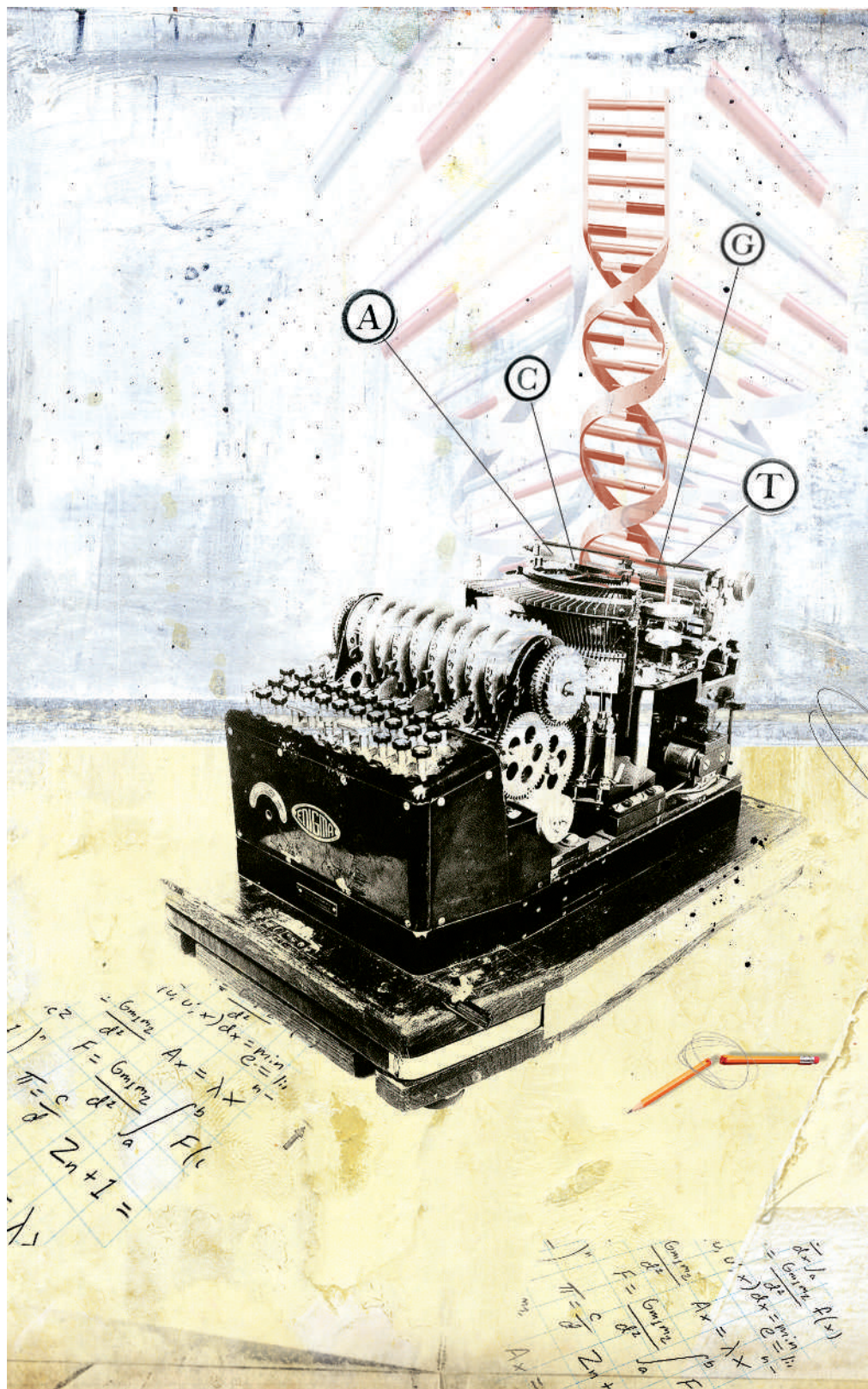
It was known that they were a little acquainted but not a syllable of real information could be procured as to what the truth was..." Reduce it to just a sequence of letters, and even a delicate phrase from Jane Austen's *Emma* becomes virtually impenetrable gobbledygook. So it was something of a triumph for Simon Shepherd when, in 2001, an algorithm he had written reconstructed all of *Emma*, word for separated word, from just such an uninterrupted string, despite being unacquainted with English vocabulary or syntax. The software worked out which groupings of letters were most likely to appear together, and thus have distinct meanings.

Shepherd, a researcher at the University of Bradford, UK, picked up much of his expertise during ten years cracking Russian codes in British Naval Intelligence. But he was not really interested in *Emma* — that was just a demonstration. His real goal was the far longer sequences of As, Gs, Cs and Ts that make up the world's genomes. Within those strings there is information that no one knows how to extract — codes that regulate, control or describe all sorts of cellular processes. And if the information is there, Shepherd thinks that number crunching should be able to pry it loose. "We are treating DNA as we used to treat problems in intelligence," he says. "We want to break the code at the most fundamental level."

That DNA contained at least one code was realized as soon as the molecule's structure was discovered. That code, cracked in the 1950s and 1960s, parses passages of DNA into three-letter combinations that correspond to particular amino acids. This is a code in the strictest sense; input determines output.

But researchers now know that there are numerous other layers of biological information in DNA, interspersed between, or superimposed on, the passages written in the triplet code. Human DNA contains tissue-specific information that instructs brain or muscle cells to produce the suite of proteins that make them brain or muscle cells. Other signals in the sequence help decide at what points DNA should coil around its scaffolds of structural proteins. These are the codes that computer buffs such as Shepherd want to crack with raw processing power — and that mainstream biologists are attacking, too, although using a rather more lab-based approach. "We need all these codes together to understand the dynamics of the cell," says computational biologist Manolis Kellis at the Massachusetts Institute of Technology in Cambridge.

The DNA sequence contains information



CODES AND ENIGMAS

There's more than one way to read a stretch of DNA, finds **Helen Pearson** — and we need to understand them all.

not just about the make-up of proteins but also about the interactions of DNA with some of those proteins, and the diverse antics of RNA. The analysis of DNA sequences is revealing patterns that have meanings at all of these levels. "Biology has probably figured out a way to squeeze every bit of information from that molecule it can," says Jason Lieb, who studies DNA-protein interactions at the University of North Carolina at Chapel Hill.

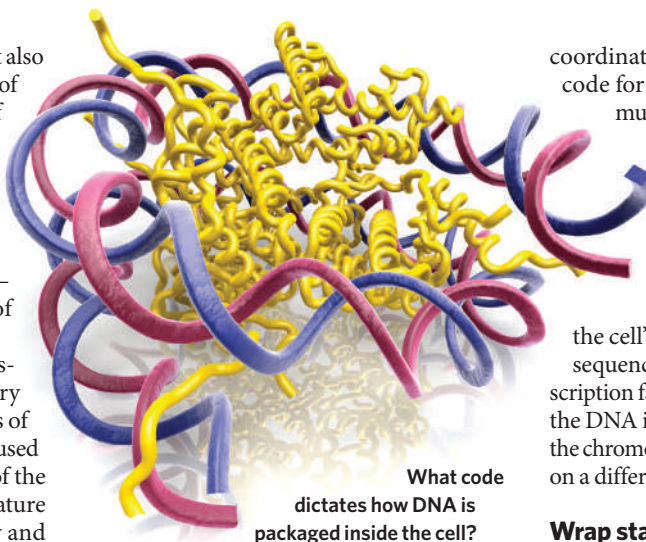
The code that is currently most exercising the minds of geneticists is the 'regulatory code' that directs the production of suites of proteins tailored to specific cell types and used at specific times. The idea is that many of the genes switched on in DNA contain signature sequences in 'promoter' regions nearby and 'enhancer' regions that may be millions of base pairs away. In a blood cell, say, these signature sequences might be bound by proteins A, B, C and D, whereas genes switched on in skin may be regulated by signature sequences that bind proteins B, C, Y and Z.

"The biggest obstacle after the sequencing of the genome has been to understand how genes are regulated and how we can see that from the sequence," says Jussi Taipale, who studies gene regulation at the University of Helsinki, Finland. "It's a more complex code than the genetic code." The first difficulty is the sheer scale of the problem. Human cells contain more than 20,000 protein-coding genes, roughly 1,500–2,000 transcription factors, which switch genes on and off, and numerous other regulatory proteins and RNAs that direct their production. The possible permutations and combinations are bewildering.

Lost in translation

One way to start solving the regulatory code en masse would be to find all the positions where each of the regulatory proteins binds within the genome. Many transcription factors show a penchant for binding specific short motifs in DNA, such as a six-letter sequence. In theory, a computer could scan for any such motifs that occur more often than might be expected by chance.

But there are drawbacks. For one thing, a given six-base-pair sequence will sometimes be a binding site and sometimes not, probably depending, in part, on whether the DNA is folded up in a way that prevents transcription factors from gaining access. For another, the way that these sites are recognized is not as specific as the binding between the bases that translate the triplet code into protein. Transcription factors recognize DNA sequences from the effects of the sequence on the outside of the helix, and although this recognition is still sequence dependent, it is not quite so precise. Some of these proteins will bind to a range of related sequences — sometimes more tightly, sometimes less so — and those subtleties of affinity, like the nuances of a social embrace, may themselves have biological meaning.



What code dictates how DNA is packaged inside the cell?

Len Pennacchio at the Lawrence Berkeley National Laboratory in California and his colleagues have begun to fathom some of these subtleties by identifying a rudimentary tissue-specific code for the human brain¹. They teased out the relevant enhancers from the human genome by comparing the human sequence to those of distant relatives such as the pufferfish (*Takifugu rubripes*), pulling out regions that didn't describe proteins but that evolution had nevertheless deemed important enough to keep intact. They then systematically inserted 167 such regions into mouse embryos and found that 45% of them provided tissue-specific ways to switch on genes.

The team identified four enhancers that boost gene activity in the developing forebrain and share several short-sequence motifs that are presumably binding sites for control proteins. By searching for similar signature sequences in the human genome, they located other forebrain enhancers, suggesting that they have found some of the sequence information that 'means' brain-specific in the regulatory code.

Taking a slightly different tack, Richard Young at the Whitehead Institute for Biomedical Studies in Cambridge, Massachusetts, and his colleagues have come up with a preliminary code that distinguishes human embryonic stem cells². They extracted human DNA bound by three key transcription factors and determined all the sequences to which those proteins chose to bind. The proteins recognize sequences near genes that need to remain active for stem cells to stay stem cells; they also recognize other sites where they seem to help shut down the genes needed for the stem cells to differentiate into other cell types. So these proteins, in combination with others, seem to stop stem cells from becoming other cell types.

Many researchers are now talking about a

coordinated effort to identify the regulatory code for all human transcription factors in multiple tissues. But they are unlikely to resolve this code without simultaneously extracting other layers of overlapping information in DNA.

A section of DNA can contain two or more layers of information that are used at different times or in different ways depending on the cell's requirements. So whether a given sequence is read as a binding site for a transcription factor to some extent depends on how the DNA involved is packaged at that point in the chromosome — and that packaging depends on a different code stored in the DNA.

Wrap stars

A human cell has to fit about two metres of DNA into a nucleus a few micrometres in diameter; that requires packing it together with proteins in a complex hierarchy of folding back and wrapping round. The fundamental element underlying all this packaging is the nucleosome — 147 base pairs of DNA wrapped around a globule of eight proteins called histones. Up to 90% of DNA is bundled up into nucleosomes, and their position influences the DNA's activity. Sequences wrapped up in nucleosomes are often less accessible to transcription factors and so less likely to be transcribed. It has been known for more than two decades that in the test-tube certain sequences are more likely to be packaged up

in nucleosomes. But in the real hustle and bustle of the cell, it was unclear to what extent such preferences get honoured.

Earlier this year, Eran Segal at the Weizmann Institute of Science in Rehovot, Israel, Jonathan Widom at Northwestern University in Evanston, Illinois, and their colleagues came the closest yet to defining a code for the position of nucleosomes³. They took DNA wrapped up in nearly 200 yeast nucleosomes, and 177 from chickens, and exposed it to enzymes that would eat up all sequences in between the

nucleosomes. They then sequenced the DNA left intact in the nucleosomes, and used computational methods to align the sequences and search for common patterns.

The team came up with a set of rules that could predict where more than 50% of nucleosomes lie in yeast and chicken DNA. "It's much less than perfect but way better than random," Widom says. The main rule is that the sequences AA, TT or TA are more likely to be found where the spiralling DNA backbone grazes the histone — they seem to help the DNA bend around the protein core.

But Segal and Widom's rules can't predict the position of a significant fraction of the



"We are treating DNA as we used to treat problems in intelligence."
— Simon Shepherd

nucleosomes. DNA's overlapping codes mean that an individual nucleosome might be usurped if regulatory proteins are already tightly bound there. The nucleosome code depends on the regulatory code, just as the regulatory code depends on the nucleosome code. In addition, the position of a nucleosome might be influenced by the way in which the nucleosome-wrapped sequence is folded and condensed yet further. "The code specifies the initial state and the cell can mess with what happens afterwards," says Oliver Rando, who studies nucleosome positioning at Harvard University.

The goal now is to find codes that govern those larger-scale features of DNA packaging, such as how the nucleosomes are twisted up into a cable of chromatin and eventually coiled into the tightly interwoven ropes of the chromosome. As yet, though, researchers have not found landmarks equivalent to nucleosomes that can guide the search for meaning — nor is it clear that they will. “There could be diffuse information spaced at hundreds of kilobases that helps package even larger pieces of the genome together,” says Lieb. “Or it could be that the exact position of those structures is not important.”

Room for manoeuvre

DNA seems well adapted to supporting a number of codes. For a start, only 1–2% of the human genome is occupied with protein-coding sequences, which leaves plenty of intervening DNA to hold other information. But many stretches of DNA in humans and other organisms manage to multitask: a sequence can code for a protein and still manage to guide the position of a nucleosome. This is possible because the triplet code is ‘degenerate’. Several slightly different triplets can code for the same amino acid, and many positions in a protein can be filled by different amino acids — so different sequences can effectively mean the same thing. This allows other signals to be imprinted on top of the first — especially when those other signals are themselves encoded with some slack.

This elegance is surely the handiwork of evolution — and if the way in which that hand had worked to solve these problems were clearer, the simultaneous decoding of all the messages involved might become easier. Perhaps ancestral organisms had simpler sequence patterns that evolution has optimized, taking advantage of its degeneracy to layer in additional information that helped organisms acquire extra complexity. Hanspeter Herzel, who specializes in statistical analyses of DNA at Humboldt University, Berlin, speculates that the space constraints of the cell may have favoured the development of nucleosomes that wound up

unruly DNA — and that their existence then encouraged the evolution of a nucleosome code in the sequence because this lowered the energetic cost of coiling up DNA. But as yet such ideas, and any help they might offer, remain tentative. “We don’t really have a phylogeny of these signals,” he says.

And in some cases, it seems that evolution may have generated patterns that have no clear biological function. In 1992, Gene Stanley at

Boston University, Massachusetts, and his co-workers created waves when they suggested that there were patterns in DNA that spanned hundreds and thousands of base pairs⁴. Stanley used the types of statistical techniques that identify correlations in climate and financial data and applied them to all the DNA sequences available in databases at the time.

Essentially, the study showed that a region with a particular chemical composition, such as one loaded with the bases A and G, is likely to be followed by a similar region hundreds or thousands of base pairs away, and that the probability of this pattern declines in a predictable way with distance. It also found that this correlation existed predominantly in DNA that did not code for protein, leading Stanley to propose that DNA previously written off as junk actually carries biological information.

The findings were controversial at the time because several other groups could not repeat aspects of the analysis, and they prompted huge interest in DNA from mathematicians and physicists. Today, these correlations are thought to be real — but interest in them has faded because, despite researchers' best efforts, the patterns have not revealed anything biologically important. Perhaps, suggests Ivo Grosse of the Leibniz Institute of Plant Genetics and Crop Plant Research in Gatersleben, Germany, the patterns could simply be traces of random evolutionary processes, such as the erosion patterns elegantly but accidentally carved into sandstone by the wind. "Long-range correlations definitely do exist, but I don't think it's some supercode imprinted in DNA," Grosse says. "We just stumbled on a feature with probably no deep biological meaning."

But to some people the thought of order with no meaning is an affront. To such minds, the idea of teasing out nature's secrets with little more than mathematical cunning and processing power will never lose its allure. When Shepherd and his graduate student Natalie Kay, in unpublished work,

ran the software that they had tried out on *Emma* over the (admittedly small) genome of Ebola virus, it identified as meaningful some sequences that, at the time, bore no annotations in genetic databases. Only later, Shepherd says, were these motifs recognized by biologists as passages that control the activity of genes or mark their ends. He thinks that approaches based on almost pure number crunching will go on to rock the field: "I firmly believe that major advances in this over the next 20, 30, 50 years will be made by the theorists, not the medics."

But researchers versed in the complexities of how DNA and proteins actually work remain convinced that their type of knowledge will remain vital to sorting the meaningful from the circumstantial. When the triplet code was first being studied, there were any number of fanciful mathematical and logical approaches to it — but the approaches that paid off were the ones informed by the greatest degree of biological insight. “Computer scientists think they can just walk in the door and solve things,” says bioinformatics expert Wyeth Wasserman at the University of British Columbia in Vancouver, Canada. “But they come to realize you need biology too.”

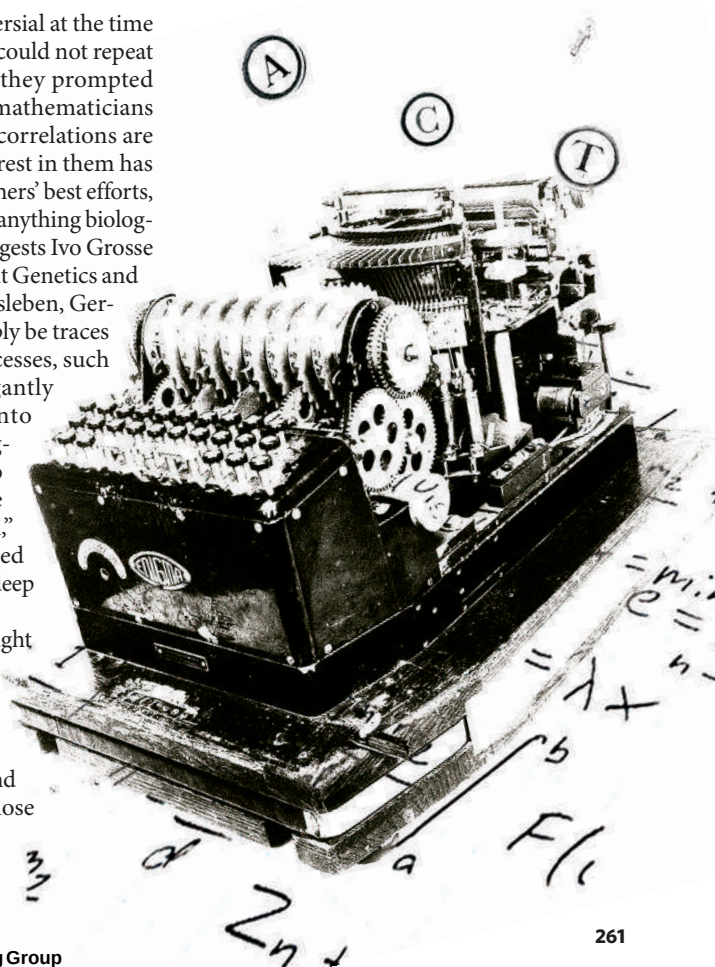
Helen Pearson is a reporter for *Nature* based in New York.

1. Pennacchio, L. A. *et al.* *Nature* doi:10.1038/nature05295 (2006).
2. Boyer, L. A. *Cell* **122**, 947–956 (2005).
3. Segal, E. *et al.* *Nature* **442**, 772–778 (2006).
4. Peng, C.-K. *et al.* *Nature* **356**, 168–170 (1992).

IS NATI CRYPTOI MUS /D ALLISON



**"It's much less
than perfect
but way better
than random."**
— Jonathan Widom



T. MATSUMOTO/AP

One man's trash...

When landfills overflow, governments need new ways to deal with garbage.

David Cyranoski visits a plant in Japan where plasma technology is turning waste into energy.



Michiaki Shigehiro has a rare problem. "I've struggled to find enough garbage," he says, in dead seriousness. He already sees a lot of waste, on average 100 tonnes per day, but he'd be far happier with twice as much.

Shigehiro is general business manager of EcoValley Utashinai, a company named after a remote city in Japan's northern island of Hokkaido. EcoValley converts heaps of refuse into energy using a plasma arc, a jolt of electricity that ionizes gas in a chamber and produces temperatures of up to 16,000 °C, or almost three times hotter than the Sun's surface. The technology is costly and must process massive amounts of trash to recoup EcoValley's ¥7-billion (US\$59-million) investment.

Utashinai city may not generate enough waste, but the rest of Japan, and the world, generally churns out too much. Every year, Japan produces about 50 million tonnes — about one kilogram per person per day — of paper, food, plastics and other garbage collectively known as municipal solid waste (MSW). The United States is almost twice as wasteful per capita, generating a total 222 million tonnes in 2005. Most of the time MSW is either stuck in a landfill or burned at a considerable cost to the taxpayer as well as releasing pollutants into the soil or air.

Americans pay between US\$30 and US\$80 per tonne, depending on the region, to dump their trash in landfills and US\$69 per tonne on average for incineration. In a small country such as Japan, where land is more scarce, local governments pay US\$200–300 per tonne to local landfills and incinerators. Landfills and incinerators also have knock-on economic effects, such as lowering property values nearby and, when improperly handled, may

create environmental hazards.

The planners behind EcoValley Utashinai thought they could solve all these trash problems at once using plasma-arc technology. In theory, waste disposal would become a profitable and environmentally benign business by converting gasified waste into energy. On paper, MSW contains one-third to one-half the energy of coal per tonne — enough to power a plant and sell excess to the national grid. But the Utashinai plasma plant is the only major energy-recycling MSW facility in operation, and it has struggled to make ends meet since opening in 2002.

Several companies, betting that they can improve on the Japanese experience, are planning their own plasma-arc facilities. Atlanta-based Geoplasma is finalizing a contract for a plant more than 10 times the size of Utashinai to be built in St Lucie, Florida, a wealthy commu-

nity where a massive landfill has lowered property values. If it succeeds, by 2009 it would be processing 2,700 tonnes of waste per day. And in September, Startech Environmental based in Wilton, Connecticut, announced a contract for a 180-tonne-per-day plant in Panama. Plasco Energy Group of Ontario is negotiating similar-sized plants in Ottawa and Barcelona.

Torching garbage

If these plants are built, it will be a vote of confidence for plasma arc, which despite its promise has not yet turned trash into gold. Those in the industry are optimistic. According to Hilburn Hillestad, Geoplasma's president for environmental affairs, the technology will catch on because of a "perfect storm of forces" — which includes increasing political and public attention on pollutants and rising energy prices.

Plasma arc is an old technology, although its use in large-scale waste disposal is new. Since the 1960s, NASA has used it to simulate the high temperatures faced by space vehicles re-entering the atmosphere. Now plasma torches are used widely for melting scrap metal or destroying hazardous materials. But it wasn't until the early 1990s that companies such as Startech and Westinghouse Plasma in Madison, Pennsylvania, which builds the plasma arcs used by Geoplasma, began developing torches for trash.

The torch is created by ionizing the air in a chamber with a powerful electric arc to generate plasma, which is then used to heat MSW, coke and limestone in a second oxygen-starved chamber. Under these conditions, the plasma torch heats the mixture to temperatures above 1,500 °C that enable it to vitrify inorganic materials in MSW without combustion occurring. The innocuous slag that results can be used as a



Ron Roberts, left, believes plants using plasma technology can remove an ugly landfill in Florida.

ST LUCIE COUNTY SOLID WASTE



Burn out: a plasma torch gasifies municipal waste, which is used to power a steam turbine at Utashinai.



turbine,” explains Osada. With a steam turbine, the gas heats water, creating steam for the turbine. In a gas turbine, the gas flow itself turns the turbine — a more efficient process, but one which potentially exposes the machinery to any corrosive gases. Indeed, Utashinai had a gas turbine from a previous project and Hitachi decided to replace it with steam turbines for fear of degradation.

Geoplasma says that, in Florida, pretreatment will reduce the amounts of corrosive gases to acceptable levels. “The scale of the St Lucie plasma plant allows for the investment in higher technology processes in order to avoid such problems,” says Hillestad.

When bidding for local government contracts, plasma companies have to compete with other waste-disposal options. Some of these also promise a measure of converting waste into energy, or lower levels of pollutants. “We researched incineration, anaerobic digestion, gasification, plasma-arc gasification, bio-reactors, and alternative landfilling methods,” says Ron Roberts, assistant solid-waste director of St Lucie county. The result: a 6,000-page report and a conclusion that there was only one proven solution to the MSW problem that would end the need for a new landfill. “The solution was plasma-arc gasification,” says Roberts.

Getting rid of the landfill was a key factor in St Lucie’s decision. Geoplasma, too, needed the landfill to guarantee it enough trash. Once the landfill is gone, the company may have to close the plant, depending on the size of the regular trash supply, but they should have recouped their installation costs by then. A lack of steady trash was one reason that Geoplasma’s plan to establish a plant in Hawaii failed. The local government would only guarantee 300 tonnes per day. “It was not enough for us to compete with rates of existing incineration plants,” says Hillestad.

Although few could argue against removing an ugly landfill, environmental groups remain wary about trace pollutants in syngas. In a 2006 report on thermal conversion strategies for MSW, California-based Greenaction for Health and Environmental Justice calls plasma-arc and other high-temperature gasification technologies “incinerators in disguise”.

The Utashinai plant says it has met all of Japan’s stringent environmental standards (although it has yet to undergo the regional government’s audit). Still Shigehiro is sceptical that the technology will catch on: “Many foreigners come to see our plant, but no one builds their own,” he says. Roberts, who has visited Utashinai, is far more upbeat. “We are not worried about it being too good to be true. We have seen the process in operation,” he says. “We certainly hope that Geoplasma has done its due diligence on the economic side of the equation.” The test of that will be St Lucie. ■
David Cyranoski is Nature’s Asian correspondent.

construction material, although it doesn’t fetch a high price.

More importantly, the heat breaks down organic molecules in the MSW. Whereas combustion generates lots of carbon dioxide, in an oxygen-limited environment the MSW is converted to a mixture of mostly carbon monoxide and hydrogen, called syngas. Syngas can be used like natural gas to power a gas turbine. Purified hydrogen could itself be used as fuel. The gas mixture is further processed to minimize the amount of other pollutants, such as nitrogen oxides and dioxins, that enter the turbine or escape into the atmosphere.

The Japanese have had some success with this technology. Utashinai’s plant pumps out 3,000 megawatts of power per year, all of which is used to run the plant. But its supply of trash is dwindling. With a population of 5,500, Utashinai holds the title of Japan’s smallest ‘city’, an appellation it earned 50 years ago when 45,000 people lived there. Rather than respond to a trash crisis, the plasma-arc plant was intended to spur the local economy by charging fees for waste disposal, which it does, and selling electricity and slag, which it does not.

On average, the plant only processes 60% of trash volume that the company expected. The facilities also suffer operational problems, though not with the plasma torch itself, and one of the two lines is often down for maintenance. When the lines are running at full capacity, there may not be enough trash. At other times, the facility turns it away.

Getting more trash from elsewhere is one option. But few citizens want their home to become a dumping ground or processing plant for trash from other areas. “No one has a good

image of garbage,” says Shigehiro. Some states in the United States will import garbage, collecting lucrative tipping fees. Big cities such as New York and Toronto export most of their waste for processing. But in Japan there is little trade in trash, local governments prefer to process waste locally than dump it elsewhere.

In Florida, St Lucie county already has a 4.3-million-tonne landfill in their backyard. Clearing that would open up 160 acres of land and increase local property values. Geoplasma will process 1,800 tonnes of new trash and 900 tonnes of landfill trash everyday, emptying the site in less than 20 years — the time Geoplasma says it will take to recoup its US\$425 million investment. Hillestad says 80% of their income will come from energy sales. The company expects to produce 160 megawatts per day, of which 120 will be sold off.

Trash trade

One difference between the St Lucie and Japanese plants that may allow higher energy output will be the use of a gas turbine, a device that turns syngas into electrical energy. The Utashinai plant uses a substantially cheaper steam turbine that can only convert about 15% of the energy to electricity. Geoplasma will use a \$40 million gas turbine that is almost 40% efficient.

This higher efficiency should help it to generate more energy from their syngas. But the Utashinai syngas contains small amounts of hydrogen chloride and sulphur oxides, says Shinichi Osada, head of the environmental systems division at Hitachi Metals, which installed the Utashinai plant. “These might attack a gas

“No one has a good image of garbage.”
— Michiaki Shigehiro

CHARMED, I'M SURE

What makes the perfect protein purification or the right reagent reaction?

Trevor Stokes investigates the weird world of good-luck lab charms.

Daniel Nelson was getting desperate. His graduate thesis at the University of Georgia, Athens, depended on his ability to isolate periodontan, an enzyme from the bacterium that causes gum disease. Yet after two years of trial and error, his protein preparation still contained too many contaminants.

Enter the lucky sombrero. For his 26th birthday, Nelson's lab mates took him to a Mexican restaurant. The birthday boy asked the waiter if he could wear a sombrero hanging on the wall — then, after sharing several pitchers of margaritas, Nelson wore the hat back to the lab. Luck finally struck: the new protein-isolation protocol he had started earlier that morning showed that he had finally isolated his favourite protein.

The 'purification sombrero' quickly gained a following among postdocs and students in the lab. "They have all heard the story and know if they don't wear the hat, they will be cursed with two years of impure protein," says Nelson, who is now an assistant professor at Rockefeller University in New York.

Similar good-luck charms populate laboratories worldwide. As childish as they may seem, such totems can benefit researchers on several levels: as a bonding activity to bring lab members together; a giver of hope when experiments are failing or a silly release to take the edge off serious research pursuits. Stuart Vyse, a psychologist at Connecticut College, New London, says superstitions can help lab members bond in a sort of "psychological hopefulness".

Magical mascots

Charms can be found in nearly every aspect of science, from biochemistry labs that depend on reagents working properly to observatories that require clear skies.

To that end, dozens of fist-sized cloth ghosts line the ceiling of the control room at the Subaru Observatory at the summit of Mauna Kea, Hawaii. In Japan, elementary-school children typically use such charms, called *teru-teru bozu* — 'sunshine monks' in English — to ensure good weather for class trips and holidays. What's good for trips is good for telescopes, say staff astronomers. "Over a period of time, each astronomer who subscribes to that superstition will add one or more of these things," says Gary Fujihara, an educator at the University of Hawaii's Institute for Astronomy at Manoa.



Toy story: an unnamed twine dragon, above, and Squeaky the dinosaur, below, assist successful PCR.

Scientists and technicians tend to pass down their charm traditions. John Gaskin, a former postdoc at Washington University in St Louis, Missouri, introduced graduate student Jason Londo to the 'amulet of power' — a fake-jewel-encrusted gold chain that students wore while conducting polymerase chain reactions (PCRs). "If it works, cool," says Londo. "And if somebody has a problem with that, we'll just see where they go when their PCR stops working." The amulet disappeared last year, but Londo replaced it with a dragon made out of twine and ribbons.

Some scientists call on their colleagues for ideas for charms, which suggests that the line between fiction and fact can be crossed in the name of social fun. On Protocol Online, a life-sciences protocol database, a recent forum read "The Gods of Molecular Biology."

— Pay homage! Share your experience!" Three authors fretted about

choosing the right sacrifice for experiments to work.

"My sacrifice of a KitKat and

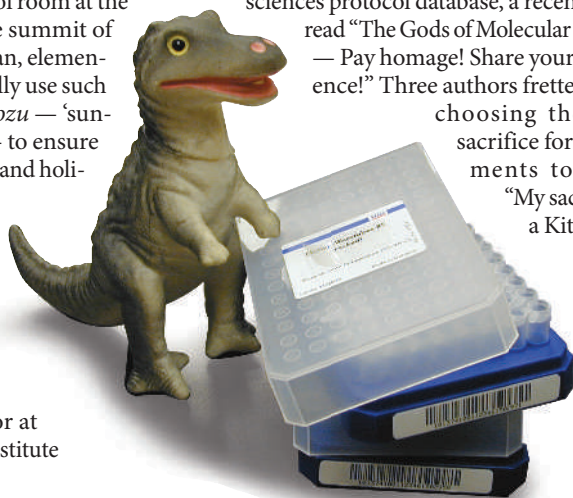
Diet Coke every morning results in having PCR reactions work (which, without the sacrifice, don't work at all)," claimed author vetticus3. Hobgoblin responded: "Our PCR spirits are very busy. They appear to us as squirrels and in our PCR room several squirrel statues and posters are placed to honour them."

Superstitious minds

Failure can also bring out the lab charms. At the University of Utah, Salt Lake City, an assistant professor of human genetics arranges voodoo candles in her lab to pay homage to the gods of cloning. "It's a little bit of a humorous comfort to people," she says, "to see that other people can make a joke out of failure." But her embarrassment over the practice — she asked not to be named, for fear of not getting tenure — suggests the charm of good-luck trinkets fizzles out as scientists become more established. Graduate students and postdoctoral fellows are more often the ones setting up good-luck charms, compared with older laboratory heads.

But Rich Whitkus, a plant evolutionary biologist at Sonoma State University in California, has kept Squeaky, a high-pitched dinosaur, for more than a decade after graduate school. "It's one of the first things that gets set up in the lab," says Whitkus. He has witnessed other researchers squeeze the dinosaur before starting PCR reactions. "Of course, we have no correlation that the number of squeaks that Squeaky gets is directly correlated with the success of PCR," he says. "We have no scientific data on that." ■

Trevor Stokes is a freelance writer in New York City.



Answering critics can add fuel to controversy

SIR — Your Editorial “To build bridges, or to burn them” and News Feature “In the name of nature” raise important points about criticism of science and how scientists should best respond (*Nature* **443**, 481 and 498–501; 2006). The News Feature concerns radical environmentalists and animal-rights activists, but the problem covers a wider area, often involving more enlightened criticism of science from outside the scientific establishment and even, sometimes, from within.

The critics feel passionately that they are right, and that their viewpoints have been unfairly neglected by the establishment. They strike a populist note. They bring into the public arena technical claims that few can properly evaluate. They are sometimes able to generate astonishing amounts of publicity. We all know examples from our own fields or from the media.

Responding to this kind of criticism can be very difficult. It is hard to answer unfair charges of élitism without sounding élitist to non-experts. A direct response may just add fuel to controversies. Critics, who are often prepared to devote immense energies to their efforts, can thrive on the resulting ‘he said, she said’ situation.

Scientists in this type of situation would do well to heed the advice in *Nature*’s Editorial. Keep doing what you are doing. And when you have the chance, try to patiently explain why what you are doing is interesting and exciting, and may even be useful one day.

Edward Witten

Institute for Advanced Study, Einstein Drive,
Princeton, New Jersey 08540, USA

Criticism: what to do about science’s bad public image?

SIR — Although you are right to state in your News Feature “In the name of nature” (*Nature* **443**, 498–501; 2006) that environmental activists such as those who fire-bombed a research facility in Olympia, Washington, are misguided, I found your focus on personal flaws and oddities to fall short. Yes, there are many vegan weirdos out there, but some of them are hard-core genetic engineers working in my laboratory. Science and anti-science have a surprising overlap in subculture. (Count the vegetarians at the next *Nature* office party.) So it’s not just subculture that is driving this small, radical and somewhat erratic movement, but a strengthening groundswell of distrust towards science and scientists.

As *Nature* is a magazine read mostly by scientists, it would be interesting to

explore and analyse how we — especially we biologists — managed to become the bad guys. How and why did our public image change from harmless geeks to state- and industry-sponsored evil-doers worthy to be a target? More importantly, what do we do about it? And how do we communicate more effectively what we are doing, why we are doing it and what the opportunities and challenges of modern science are?

So I very much look forward to a follow-up article that won’t just have me worried that the friendly tofu-lover I see in one of the Athens clubs tonight will burn down my lab, but will stir me to engage the public in a more effective way.

Boris Striepen

Center for Tropical and Emerging Global Diseases,
Paul D. Coverdell Center, University of Georgia,
Athens, Georgia 30602, USA

Creationists attack secular education in Russia

SIR — Evolutionary biology has come under attack from creationists in the United States and in several European countries, most recently, Poland (“Polish scientists fight creationism” *Nature* **443**, 890–891; 2006). Creationists in Russia are also attacking darwinism, and indirectly attacking the principle of a scientifically founded, secular education system.

The Russian case concerns 15-year-old Maria Schreiber and her family, who have filed a complaint to the federal court in St Petersburg demanding a “free choice” for the girl, as her religious sensibilities have been hurt by “Darwin’s controversial hypothesis” (reported in *Gazeta.ru* 27 October 2006). The plaintiffs criticize the biology textbook for classes 10–11 and wish to change it. The court case began on 25 October and may be decided by mid-December.

In Poland, as your News story notes, the creationist case is supported by the minister and deputy minister of education, and the case against evolution has supporters among prominent politicians. In Russia, Schreiber’s case has support from the powerful Orthodox church, and the family’s lawyers are distant relatives of the last Russian tsar.

As historians of biology and evolutionary biologists, we are aware of the strong tradition in evolutionary biology in Russia, where prominent scholars did important work throughout the twentieth century, particularly in pioneering research into the population genetics of natural populations.

Maybe we are now seeing the delayed effects of 70 years of enforced atheism and official support for darwinism in the Soviet Union, which kept creationism at bay until its collapse in 1991. With the religious freedom that has been allowed since then, numerous

churches have become active in Russia: there were 21,664 religious organizations registered by the Ministry of Justice in 2004. Some of them support creationism, and they find an audience that may relate darwinism to Soviet ideology rather than to empirical natural science.

Georgy S. Levit*, Uwe Hoffeld*, Lennart Olsson†

*Ernst-Haeckel-Haus/AG Biologiedidaktik,
Friedrich-Schiller-Universität Jena, Berggasse 7,
D- 07745 Jena, Germany

†Institut für Spezielle Zoologie und
Evolutionsbiologie mit Phyletischem Museum,
Friedrich-Schiller-Universität Jena, Erbertstraße 1,
D- 07743 Jena, Germany

Creationism, evolution: nothing has been proved

SIR — In your News story “Polish scientists fight creationism” (*Nature* **443**, 890–891; 2006), you incorrectly state that I have called for the “inclusion of creationism in Polish biology curricula”. As well as being a member of the European Parliament, I am a scientist — a population geneticist with a degree from Oxford University and a PhD from the University of Toronto — and I am critical of the theory of evolution as a scientist, with no religious connotation. It is the media that prefer to consider my comments as religiously inspired, rather than to report my stated position accurately.

I believe that, as a result of media bias, there seems to be total ignorance of new scientific evidence against the theory of evolution. Such evidence includes race formation (microevolution), which is not a small step in macroevolution because it is a step towards a reduction of genetic information and not towards its increase. It also includes formation of geological strata sideways rather than vertically, archaeological and palaeontological evidence that dinosaurs coexisted with humans, a major worldwide catastrophe in historical times, and so on.

We know that information exists in biology, and is transferred over generations through the DNA/RNA/protein system. We do not know its origin, but we know it exists, can be spoiled by mutations, but never improves itself spontaneously. No positive mutations have ever been demonstrated — adaptations to antibiotics or herbicides are equivalent to immunological adaptation to diseases, and not a creation of a new function.

We keep on searching for natural explanations of everything in nature. If we have no explanations we should say so, and not claim that an unproven theory is a fact.

Maciej Giertych

Institute of Dendrology,
Polish Academy of Sciences,
62-035 Kórnik, Poland

COMMENTARY

Safe handling of nanotechnology

The pursuit of responsible nanotechnologies can be tackled through a series of grand challenges, argue **Andrew D. Maynard** and his co-authors.

When the physicist and Nobel laureate Richard Feynman challenged the science community to think small in his 1959 lecture ‘There’s Plenty of Room at the Bottom’, he planted the seeds of a new era in science and technology. Nanotechnology, which is about controlling matter at near-atomic scales to produce unique or enhanced materials, products and devices, is now maturing rapidly with more than 300 claimed nanotechnology products already on the market¹. Yet concerns have been raised that the very properties of nanostructured materials that make them so attractive could potentially lead to unforeseen health or environmental hazards².

The spectre of possible harm — whether real or imagined — is threatening to slow the development of nanotechnology unless sound, independent and authoritative information is developed on what the risks are, and how to avoid them³. In what may be unprecedented pre-emptive action in the face of a new technology, governments, industries and research organizations around the world are beginning to address how the benefits of emerging nanotechnologies can be realized while minimizing potential risks⁴. Yet despite a clear commitment to support risk-focused research, opportunities to establish collaborative, integrated and targeted research programmes are being missed⁵. In September, Sherwood Boehlert, chair of the US House Science Committee, commented in a hearing that “we’re on the right path to dealing with the problem, but we’re sauntering down it when a sense of urgency is required”. And in October, Britain’s Royal Society raised concerns that the UK government had not made enough progress on reducing the uncertainties surrounding the health and environmental impacts of nanomaterials⁶.

The risks

As research leaders in our respective fields, we recognize that systematic risk research is needed if emerging nano-industries are to thrive. We cannot set the international research agenda on our own, but we can inspire the scientific community — including government, industry, academia and other stakeholders — to move in the right direction. So we propose five

“Understanding and preventing risk often has a low priority in the competitive world of research funding.”

grand challenges to stimulate research that is imaginative, innovative and above all relevant to the safety of nanotechnology.

Fears over the possible dangers of some nanotechnologies may be exaggerated, but they are not necessarily unfounded. Recent studies examining the toxicity of engineered nanomaterials in cell cultures and animals have shown that size, surface area, surface chemistry, solubility and possibly shape all play a role in determining the potential for engineered nanomaterials to cause harm⁷. This is not surprising: we have known for many years that inhaled dusts cause disease, and that their harmfulness depends on both what they are made of and their physical nature. For instance, small particles of inhaled quartz lead to lung damage and the potential development of progressive lung disease, yet the same particles with a thin coating of clay are less harmful⁸. Asbestos presents a far more dramatic example: thin, long fibres of the material can lead to lung disease if inhaled, but grind the fibres down to shorter particles with the same chemical make-up and the harmfulness is significantly reduced⁹.

It is generally accepted that, in principle, some nanomaterials may have the potential to

cause harm to people and the environment. But the way science is done is often ill-equipped to address novel risks associated with emerging technologies. Research into understanding and preventing risk often has a low priority in the competitive worlds of intellectual property, research funding and technology development. And yet there is much at stake in how potential nano-specific risks are understood and managed. Without strategic and targeted risk research, people producing and using nanomaterials could develop unanticipated illness arising from their exposure; public confidence in nanotechnologies could be reduced through real or perceived dangers; and fears of litigation may make nanotechnologies less attractive to investors and the insurance industry.

The science community needs to act now if strategic research is to support sustainable nanotechnologies, in which risks are minimized and benefits maximized. Our five grand challenges are chosen to stimulate such research, as well as bring focus to a range of complex multidisciplinary issues. The challenges span the next 15 years, and their successful achievement will depend on coordination, collaboration, resources and ingenuity. They are not comprehensive — there is essential research that is not covered here — but they do form a framework on which others can build.



D. RAMSEY

Potential health risks from exposure to engineered nanomaterials must be understood and minimized.

The challenges

Develop instruments to assess exposure to engineered nanomaterials in air and water, within the next 3–10 years.

Because nanotechnologies are diverse and exposures to nanomaterials will vary widely, assessing exposure and potential impacts on health or the environment will require multiple sensor types operating under different conditions. Three issues stand out as fertile ground for innovative research: monitors for airborne exposure, detectors for waterborne nanomaterials, and smart sensors that can measure both exposure and potential hazards.

We don't yet know which aspects of airborne nanomaterials should be measured — number, surface area or mass concentration, a combination of these, or something else entirely. But people working with nanomaterials urgently need inexpensive personal aerosol samplers that are capable of measuring exposure in the workplace and environment. This universal aerosol sampler would log exposure against aerosol number, surface area and mass concentration simultaneously, and provide a historic record that can be interpreted in the light of new knowledge and new exposure monitoring paradigms. It would be portable, sufficiently inexpensive to ensure widespread use, and available commercially within the next 3 years.

Effluent from nanomanufacturing processes, use of nanoparticle-containing substances such as sunscreens, and disposal of nanomaterial-containing products, will inevitably lead to increasing quantities of engineered nanomaterials in water systems. If we cannot track these materials, it will be almost impossible to determine how benign or harmful their presence is. The second challenge therefore is to develop

instruments that can track the release, concentration and transformation of engineered nanomaterials in water systems (including liquid-based nanotechnology consumer products), within the next 5 years.

Advances in information technology and sensor design are leading to the development of smart sensors that combine information on various aspects of exposure and hazard in a way that is useful for decision-making. The concept is embodied in radiation monitors and biomonitoring, but has not yet been extended to engineered nanomaterials. The final part of this challenge therefore is to develop smart sensors that indicate potential harm to human health. An example would be sensors that simultaneously detect airborne nanoparticles and determine their potential to generate reactive oxygen species — possibly providing early indications of harm. Such sensors should be available within the next 10 years.

Develop and validate methods to evaluate the toxicity of engineered nanomaterials, within the next 5–15 years.

There are many aspects to evaluating nanomaterial toxicity. But there are three that we consider crucial for stimulating high-quality research and preventing the unnecessary use of hazardous nanomaterials: validated screening tests, developing viable alternatives to *in vivo* tests, and determining the toxicity of fibre-shaped nanoparticles.

The enormous diversity of engineered nanomaterials with different sizes, shapes, compositions and coatings matches, and possibly exceeds, that of conventional chemicals. High-throughput protocols that are benchmarked and validated are urgently needed to screen for potential hazards. The first part of this

challenge is to reach international agreement on a battery of *in vitro* screening tests for human and environmental toxicity within the next 2 years, and to validate these tests within the next 5 years. Essential to this challenge will be the widespread and global availability of standard nanoparticle samples to allow comparison and refinement of methods across government, industry and academic laboratories.

Although testing *in vivo* will continue to provide the most relevant information on human (and other organism) hazards, there is an economic and ethical impetus to minimize the burden of animal testing. Emerging technologies — including nanotechnology — are providing new possibilities for simulating and predicting nanomaterial behaviour in living organisms. We propose that relevant and validated alternatives to *in vivo* toxicity testing of engineered nanomaterials be developed over the next 15 years.

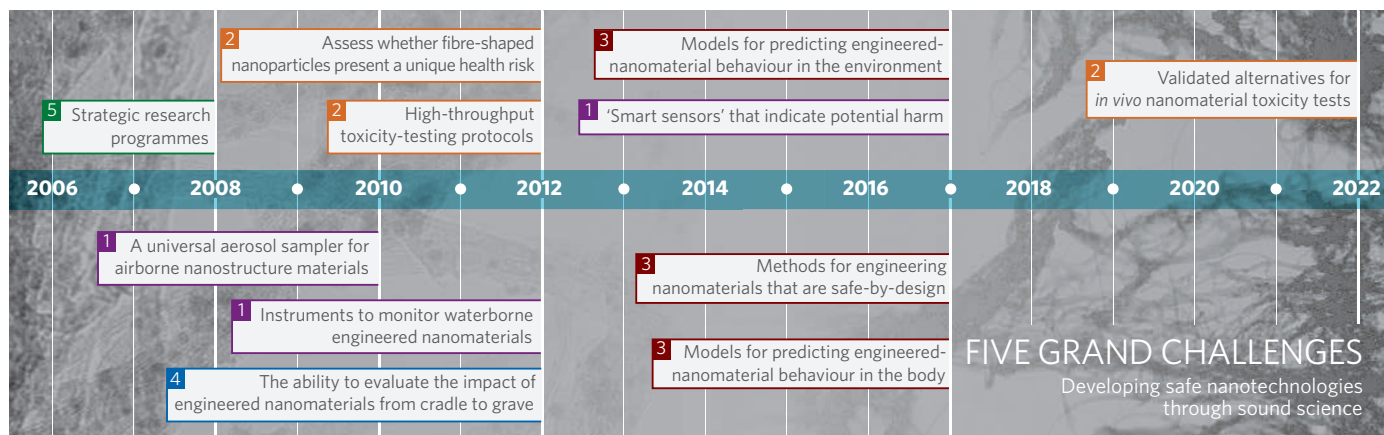
Fibre-shaped nanomaterials possibly represent a unique inhalation hazard, and their pulmonary toxicity should be evaluated as a matter of urgency. Inhalation of a sufficient dose of asbestos fibres can lead to the malignant disease mesothelioma, the causation of which is related to the length, width and chemistry of the fibres, as well as their ability to persist in the lungs.

Although it is not clear whether fibre-shaped nanoscale particles formed from carbon and other materials will behave like asbestos or not, some materials are sufficiently similar to cause concern: any failure to pick up asbestos-like behaviour as early as possible would be potentially devastating to the health of exposed people and to the future of the nanotechnology industry. We propose that the potential health impact of high-aspect-ratio, biopersistent engineered nanotubes, nanowires and nanofibres is systematically investigated within the next 5 years.

"Without strategic risk research, public confidence in nanotechnologies could be reduced through real or perceived dangers."



Nanotechnology is rapidly advancing, with more than 300 nanoproducts already on the market.



Develop models for predicting the potential impact of engineered nanomaterials on the environment and human health, within the next 10 years.

To assess the safety of complex multicomponent and multifunctional nanomaterials, scientists will need systems capable of predicting the potential impact of new nanomaterials, devices and products. Once again our challenge here has three parts. First, to develop validated models capable of predicting the release, transport, transformation, accumulation and uptake of engineered nanomaterials in the environment. In parallel, validated models must be developed that are capable of predicting the behaviour of engineered nanomaterials in the body, including dose, transport, clearance, accumulation, transformation and response. These models should: relate the physical and chemical characteristics of nanomaterials to their behaviour; allow an integrated approach to predicting potential impact of engineered nanomaterials and nanoproducts; and estimate impact within susceptible populations.

Third, to use predictive models for engineering nanomaterials that are safe by design. This might include engineering the nanomaterials in ways that enhance desired properties while suppressing hazardous ones, or creating fail-safe mechanisms that ensure a transition to benign materials upon disposal.

Develop robust systems for evaluating the health and environmental impact of engineered nanomaterials over their entire life, within the next 5 years. Thinking in terms of life cycles leads to a holistic approach to managing risks and benefits. Developing robust ways of evaluating the potential impact — good or bad — of a nanoproduct from its initial manufacture, through its use, to its ultimate disposal will stretch both scientific and policy communities, but will lead to new methodologies that are widely applicable.

Develop strategic programmes that enable relevant risk-focused research, within the next 12 months. Ultimately, systematic and

organized risk research will empower industry, consumers and policy-makers to make the best decisions about the development and application of emerging nanotechnologies. As end-users of the scientific data, these communities must play a central role in shaping what is done and how. Government research strategies that systematically reduce uncertainty surrounding the potential impact of nanotechnologies and support science-based oversight are essential to the safe development of nanotechnology. But these must be complemented by, and integrated with, industry-led research. We highlight three areas that we believe are critical to the success of such risk research: col-

laboration, communication and coordination.

The first challenge is identifying mechanisms that enable collaborative research programmes — whether interdisciplinary, between government and industry, or between different stakeholders. Virtual interdisciplinary research centres and networks are one way of stimulating collaboration, as long as they are accompanied by adequate resources. We would also encourage joint government–industry partnerships that underpin good product stewardship and oversight, while being transparent and credible.

Communicating research on nanotechnology risks and benefits outside the scientific community is challenging, but is essential for a risk dialogue based on sound science. This means developing communication activities that enable technical information to be summarized, critiqued and ultimately synthesized for various interested parties, including decision-makers and consumers. The advent of the Internet provides an ideal venue for such activities and we encourage its use in communicating with the end-users of risk-based science.

Finally, a global understanding of nanotechnology-specific risks is essential if large and small industries are to operate on a level playing field, and developing economies are not to be denied essential information on designing safe nanotechnologies. We propose that mechanisms, networks and meetings are established

that enable international information-sharing and coordination between the public and private sectors.

Nanotechnology comes at an opportune time in the history of risk research. We have cautionary examples from genetically modified organisms and asbestos industries that motivate a real interest, from all stakeholders, to prevent, manage and reduce risk proactively. We have a global research infrastructure that supports international collaboration. We also have revolutions in biotechnology, sensing and computation that are transforming how health-focused research is performed. If the global research community can take advantage of these circumstances and rise to the challenges we have set, then we can surely look forward to the advent of safe nanotechnologies. ■

Andrew D. Maynard is at the Woodrow Wilson International Center for Scholars; his co-authors are Robert J. Aitken, Tilman Butz, Vicki Colvin, Ken Donaldson, Günter Oberdörster, Martin A. Philbert, John Ryan, Anthony Seaton, Vicki Stone, Sally S. Tinkle, Lang Tran, Nigel J. Walker and David B. Warheit
e-mail: Andrew.Maynard@wilsoncenter.org.

1. *The Nanotechnology Consumer Products Inventory* (Woodrow Wilson International Center for Scholars, Washington DC, 2006). Published online at www.nanotechproject.org/consumerproducts.
2. *Nanoscience and Nanotechnologies: Opportunities and Uncertainties* (The Royal Society and The Royal Academy of Engineering, London, 2004).
3. *Taking Action on Nanotech Environmental, Health and Safety Risks* (Lux Research, New York, 2006).
4. *Report of the OECD Workshop on the Safety of Manufactured Nanomaterials: Building Co-operation, Co-ordination and Communication* (Organization for Economic Co-operation and Development, Paris, 2006).
5. Maynard, A. D. *Nanotechnology: A Research Strategy for Addressing Risk* (Woodrow Wilson International Center for Scholars, Washington DC, 2006).
6. *Two-Year Review of Progress on Government Actions: Joint Academies' Response to the Council for Science and Technology's Call for Evidence* (The Royal Society and The Royal Academy of Engineering, London, 2006).
7. Oberdörster, G., Oberdörster, E. & Oberdörster, J. *Environ. Health Perspect.* **13**, 823–840 (2005).
8. Donaldson, K. & Borm, P. J. A. *Ann. Occup. Hyg.* **42**, 287–294 (1998).
9. Davis, J. G. et al. *Br. J. Exp. Pathol.* **67**, 415–430 (1986).

Disclaimer: The opinions in this paper reflect those of the authors and do not necessarily represent the official views or policies of their affiliations.

BOOKS & ARTS

Opposition to science

A molecular biologist explores the gulf between spirituality and his own rationalist viewpoint.

Challenging Nature: The Clash of Science and Spirituality at the New Frontiers of Life
by Lee M. Silver

HarperCollins: 2006. 464 pp. \$26.95

James T. Bradley

In his essay *The Two Cultures* (Cambridge University Press, 1959), C. P. Snow detailed the gulf of incomprehension separating 'literary intellectuals' from scientists. Lee Silver's new book, *Challenging Nature*, although not mentioning Snow's work, focuses on a similar disconnect: between rationalist/physicalist and spiritual world views. The book's engaging prose informs and, above all, provokes.

Silver is an internationally renowned molecular biologist and self-styled rationalist, and acknowledges his membership of a scientific community "in which life is assumed to be combinations of complex molecules and information flow between them". His exploration of spirituality took him across five continents, where he learnt, through conversation and first-hand observation, about common people's "beliefs about life and spirits, and their place in the universe". His descriptions of these beliefs are fascinating, and range from cremation and skull-breaking ceremonies for releasing the deceased's spirit, to disparate beliefs about when developing humans acquire souls or special moral status.

His interest in spirituality springs from his attempt to understand the basis for the anxiety, anger and fear provoked by modern biotechnologies. This resistance comes from opposite poles of the political spectrum — the religious right, especially in the United States, which believes we should not 'play God' by manipulating genomes and embryos, and the environmentalist left, which sanctifies nature. But before analysing this opposition, he discusses the meaning of science, faith, religion and spirituality.

Modern science is based on a cyclic interplay between empirical and theoretical research. Scientific theories require falsifiability, and empirical findings are often probabilistic, never absolute. By contrast, faith is "a belief held about an aspect of oneself or the larger world in the absence of factual evidence to support it... usually absolute rather than probabilistic, and often not falsifiable."

Religion resists concise definition. Silver discusses a spectrum of faith-based beliefs in a transcendent god, from deism to fundamen-



Playing God or defiling nature? Biotechnology is opposed by both ends of the US political spectrum.

talism. He points out that some fundamentalists claim revelation rather than faith as the basis for absolute truth, and cites a 2003 Harris poll showing that 90% of Americans believe in God and 84% in an afterlife. The truth, according to Silver, is that "modern science — as a body of knowledge, as a profession, and as a means for achieving intellectual stature — shows a striking correlation with skepticism or outright rejection of the religious and spiritual beliefs held by the vast majority of nonscientists." His critique of spirituality suggests that the nearly universal, cross-cultural belief in a transcendent reality is a form of emotionalism, a product of natural selection with an underlying biology that, in overdose, can tip one towards psychosis, schizophrenia or other delusional experiences.

Silver is relentless as he seeks to expose the spiritual (non-rational) bases for both anti-biotechnology and pro-Mother Earth positions. His critique spares no views against the use of embryonic stem cells, cloning, genetic engineering and other biotechnologies, or for ecosystem preservation and homeopathic medicine, unless they are underpinned by science-based arguments. He does not claim

that spirituality is all bad, or that all biotechnological applications are good and risk-free, but some well-developed examples would have been welcome to back up this claimed openness. The crucial question is how, or whether, spirituality should contribute to decision-making about our biotechnological future.

Inevitably, writes Silver, "human nature will remake all of Mother Nature in the image of the idealized world that exists within our own minds." He suggests that human-directed evolutionary processes and designer ecosystems cannot be any worse than "the vagaries of Mother Nature who cares not at all for aesthetics, humankind, or the survival of anything in particular". In my opinion, this view underestimates the 'wisdom' of natural selection in structuring stable, integrated ecosystems, and manifests the creative loss that Snow laments when two cultures fail to communicate: "The clashing point" of two cultures "ought to produce creative chances... But they are... in a vacuum because those in the two cultures can't talk to each other."

Challenging Nature lays out the incredible richness in two ways of knowing, but I sense no lament over the ocean of incomprehension

ZUMA PRESS/NEWS.COM

separating them. Can the gulf be bridged? One example of creativity at the interface of rationalism and spirituality in the biological realm is conservation biologist Aldo Leopold's *A Sand County Almanac* (Oxford University Press, 1949). Leopold gives three reasons for preserving native wilderness areas: science, wildlife and recreation. On the value of preserving wilderness for the few who practise the primitive arts of canoeing and packing, Leopold wrote: "Either you know it in your bones, or you are very, very old." And on recognizing the cultural value of wilderness, he wrote that it is "a question of intellectual humility". I believe

Silver would view these as spirituality-based statements, yet we could do worse than accept Leopold's wisdom and the creatively combined rationalism and spiritualism informing it.

Most left-brained people will love this book. It may annoy right-brained people, but their response to it will enhance the creative, democratic dialogue so badly needed on the issues addressed. ■

James T. Bradley is in the Department of Biological Sciences, Auburn University, Auburn, Alabama 36849, USA. He is currently writing a book for non-scientists called *Twenty-First Century Biotechnologies and Human Values*.

Biology's big idea

In the Beat of a Heart: Life, Energy, and the Unity of Nature

by John Whitfield

Joseph Henry Press: 2006. 261 pp. \$27.95

David Robinson

D'Arcy Wentworth Thompson is the hero with whom John Whitfield begins and ends his engaging book, *In the Beat of a Heart*. Teaching "at a provincial university in a coarse, industrial Scottish city" in the early twentieth century, Thompson's obsession was the search for principles to unify the diversity of life. This is the springboard for Whitfield's lively account of more recent attempts to answer questions that Thompson posed.

Thompson's 1917 book *On Growth and Form* famously depicted his (often incorrect) ideas about how organisms are as much the products of physics as of natural selection. A polymath of astonishing accomplishment, Thompson was better equipped than most to appreciate how physical simplifications of nature can reveal things about the living world that traditional approaches cannot uncover. He believed that however much evolution causes animals or plants to vary in delightful ways, feathers and foliage hide universal features of structure or function that reflect unbreakable physical laws. Identifying those features was Thompson's goal. His work needed bold generalizations, and he was unafraid to look at nature in a different way from everyone else. Eighty years after *On Growth and Form* first appeared, its unfashionable philosophy was emulated by the main protagonists of Whitfield's book.

In 1997, physicist Geoffrey West and two ecologists, Jim Brown and Brian Enquist, developed a theory to explain why many familiar biological patterns vary as quarter-powers of body mass. For example, a mammal's heart rate varies, on average, as its body mass to the power $-1/4$; an animal's lifespan varies as its body mass to the power $1/4$; tree height varies as body mass to the power $1/4$, but tree density in a forest varies as body mass to the power $-3/4$, and so on. These patterns suggest that



Did Jonathan Swift use quarter-power scaling to decide Gulliver's food intake in *Gulliver's Travels*?

there is an underlying order to the living world, but how could such order possibly arise among organisms and processes so diverse?

West, Brown and Enquist answered this question by explaining why metabolic rate tends to scale as body mass to the power $3/4$, one of the most fundamental and enigmatic of biological relationships (and, Whitfield tells us in passing, one that was implied in *Gulliver's Travels*). With thompsonian economy and elegance, West and his colleagues specified the kind of branched vessels needed to transport blood efficiently around an idealized organism, worked out how those vessels could be packed optimally into bodies of different sizes, and predicted how the organism's metabolic rate would then vary with its mass. The resulting algebra yielded the magic number $3/4$. From this initial triumph, the theory has since evolved spectacularly to account for many broad features of metabolism, ecosystem processes, life histories,

developmental rates, community structure, the global carbon cycle, tumour growth and so on, and, somewhat improbably, even makes predictions about human fertility and the wealth of nations. Like evolution by natural selection and the DNA double helix, this theory explains so much with so little. It is breathtaking in its ambition and scope.

Any new theory that is apparently so omniscient will attract as many grumbles of doubt as gasps of admiration, and this one is no exception. Its fans accept as strengths its physical simplifications, its neglect of biological detail and its mathematical reasoning, all of which leave critics uncomfortable. But for many, the clincher is this: within the limits of its assumptions and the bending of its rules by biological variation, West, Brown and Enquist's theory accurately predicts an extraordinary range of phenomena. No comparable idea yet matches it, despite its inevitable limitations.

Whitfield does a fine job of describing the logic behind the theory and its antecedents. He unpacks its key assumptions and describes what the fractal plumbing system responsible for quarter-power scaling would look like. No armchair pundit, Whitfield interviewed the theory's authors and their colleagues, censured trees in Costa Rican forests with Enquist's team of students and postdocs, and spent a few less arduous hours having his own metabolism measured in London. His first-hand experiences at the subject's coalface are vividly readable. Whitfield's later chapters consider how metabolism relates to biodiversity and biogeography, and how it might dovetail with genetics. They also dwell on how these grand ideas might apply, or not, to the largest part of the tree of life: microbes. Overall, Whitfield's book provides the best available introduction to West, Brown and Enquist's big idea.

But is the big idea correct and so universally applicable? Whitfield does not ignore its critics, but they get relatively thin coverage despite their prominence in the pages of specialist ecology journals. This is understandable in a book of this type, which sets out to popularize as much as inform, but it implies that the theory itself is virtually home and dry. The most explicit cautionary note comes from West himself: "If it's wrong, it's wrong in some really subtle way."

West and his colleagues have been almost as vigorous in defending their idea as they have in using it to attack ever more diverse biological problems, and strong personalities on both sides of the debate have generated robust exchanges. To his credit, Whitfield resists the temptation to overdramatize the disputes that often accompany important scientific developments. Instead he focuses on the power of a beguilingly simple idea about how the living world might work, and on the remarkable men who conceived it. ■

David Robinson is at the School of Biological Sciences, University of Aberdeen, Aberdeen AB24 3UU, UK.

A complex mind

Il Nobel dimenticato: La vita e la scienza di Camillo Golgi (The Forgotten Nobel: the Life and Science of Camillo Golgi)

by Paolo Mazzarello

Bollati Boringhieri: 2006. 660 pp. €40

www.bollatiboringhieri.it

Elio Raviola

I can hardly suppress my irritation when, in reading a paper, I come across the term 'golgi complex' with a lower-case 'g'. This is because, as a medical student at the University of Pavia in Italy, I grew up among the epigones of the 'Golgi school', and as an intern there I learned the tricks of Camillo Golgi's 'black reaction' for staining neurons directly from A. Pensa, one of his surviving students. However, I was shocked that people in the lab still uncritically believed in the existence of Golgi's 'diffuse nervous network', rejected the well-established fact that neurons are cells independent of one another, and ignored the works of Charles Sherrington and Bernard Katz on synapses.

What was it in Golgi's personality that commanded such blind faith that his mistakes were carried over to the pupils of his pupils? Why am I still upset that, in the mind of most English-speaking neuroscientists, Golgi's name is indelibly associated with his staunch defence of the flawed reticularist theory? Is it simply that a few wrong interpretations should be allowed to taint his reputation despite an otherwise stellar scientific career? More likely, I am also under the spell of Golgi's enigmatic personality. After all, he is the genius who revolutionized neuroanatomy with a method that, for the first time, revealed neurons in their entirety and with the outmost clarity. He was the first to unequivocally describe dendrites and axons, to distinguish interneurons from relay neurons, and to provide descriptions of the cerebellar cortex, olfactory bulb and hippocampus. He was the cytologist who identified the apparatus that bears his name (the only eponym that today defines a major component of the cell). He was the general pathologist who discovered the pathophysiology of malaria, linking its febrile episodes to the completion of the protozoan's division within the erythrocytes and the release of the resulting merozoites into the bloodstream — he should have received a second Nobel prize for this discovery alone. And these are just a few of a stream of findings, each sufficient to represent a lifetime of achievement for an ordinary scientist.

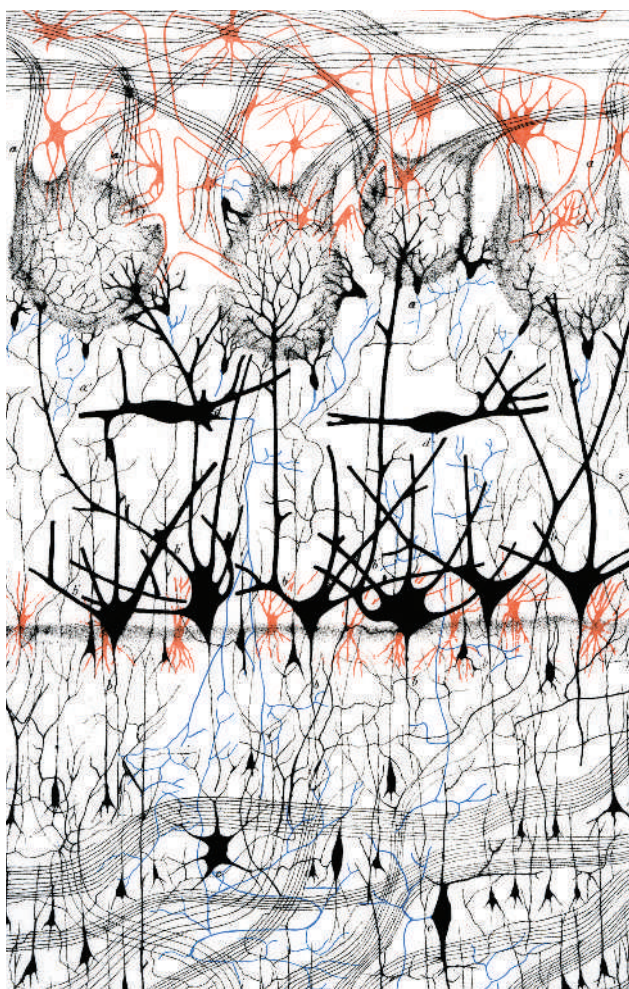
The answers to all these questions, and many more, are contained in *Il Nobel dimenticato*, a marvellous book by Paolo Mazzarello, historian of medicine at the University of Pavia. Published to mark the centenary of the Nobel prize awarded to Golgi and Ramon y Cajal, the book represents the most complete analysis yet of the life, personality, discoveries and ideas of Golgi. It comes seven years after Mazzarello's first work on Golgi, *The Hidden Structure* (Oxford University Press, 1999). Mazzarello is clearly partial to his subject, as one might expect of a historian who was a student at the University of Pavia and who still walks every day past the sombre, imposing statue of Golgi in the university's courtyard. Being at Pavia

he viciously attacked his competitor's ideas and research style.

Mazzarello provides a thorough psychological portrait of this introverted, shy and intimidating man, who liked to quote the Roman historian Tacitus: "I never regretted being silent, but always regretted having talked." Pensa told me that as an intern in Golgi's laboratory, he was paralysed by the "icy stare of his cerulean eyes". In stark contrast to the ebullient, passionate Cajal, Golgi was an uninspiring speaker who read from his notes, leading many students, including some of the interns at his laboratory, to desert his lectures; he had to threaten them with expulsion to make them sit his classes. Mazzarello deftly discusses why such a skilled microscopist and experimental scientist, a committed positivist who rejected any metaphysics, continued to defend the lost cause of the reticularist theory. However, his prestige, political acumen and careful judgement (his mantra was 'educative caution') helped him become president of the university, a city councillor, a member of the Italian Congress and a prime mover in the cut-throat competitions for departmental chairs in Italian medical faculties.

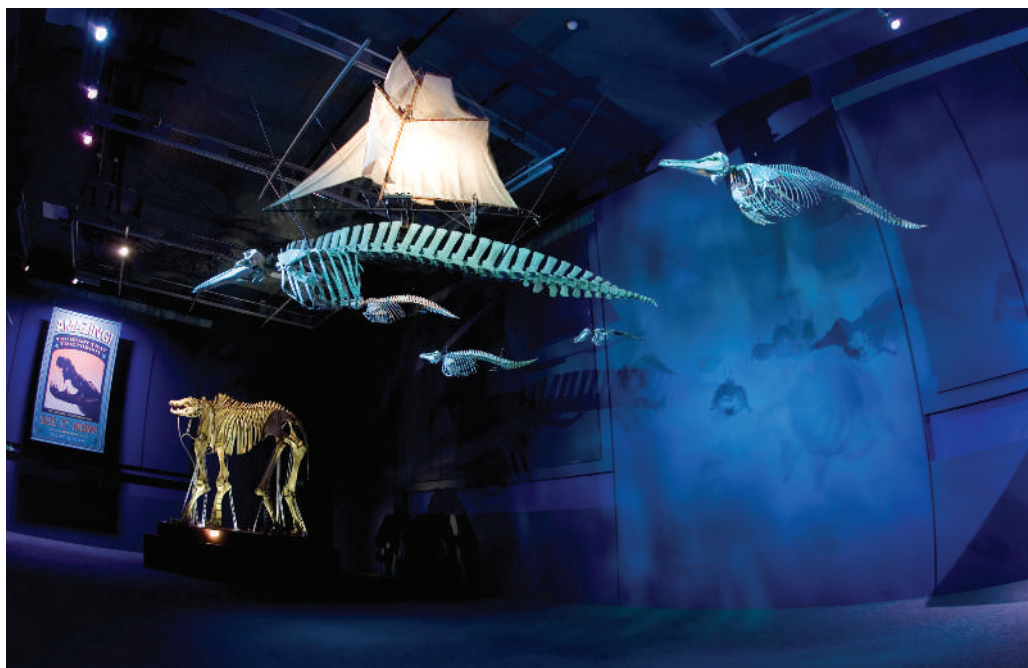
The book also provides an accurate description of academic life at the University of Pavia, long after the glorious days of Lazzaro Spallanzani, Alessandro Volta and Antonio Scarpa, when it had been second only to Vienna as the seat of learning in the Austro-Hungarian empire. It tells of the political affairs in Pavia, a small, provincial city in northern Italy, which had just become part of a unified, independent kingdom. And it provides a lively account of the interactions among biomedical scientists in Europe in the second half of the nineteenth century, an era of great discoveries, friendships and enmities, violent personal attacks and romantic, rhetorical praise.

I hope Mazzarello's book will be translated into English. Perhaps a better knowledge of Golgi's extraordinary discoveries, and a perusal of Mazzarello's arguments, might persuade the members of the Cajal Club — a society created in 1947 by US neuroanatomists in honour of 'el sabio de Madrid' — to change its name to the Cajal and Golgi Club. After all, as Mazzarello points out, the discovery of the black reaction came out of the blue, as a total surprise, and so represents an almost unique epistemological event in the history of biomedical science. And without the black reaction, Cajal may not have turned his attention to the nervous system in the first place. ■ Elio Raviola is in the Department of Neurobiology, Harvard Medical School, Boston, Massachusetts 02115, USA.



The 'black reaction' allowed Golgi to stain and visualize neurons.

gives him access to Golgi's original slides and allows him to handle his beautiful drawings in the university's museum. Even so, the book contains the most thorough and dispassionate analysis of the controversy between Golgi and Cajal, a convincing exposition of the prevailing *zeitgeist* at the time of their discoveries, and a justification (or lack thereof) of Golgi's behaviour at the Nobel prize ceremony, where



Flight of fancy? Michael McMillen's exhibit *The Flying Dutchman* playfully combines suspended porpoise skeletons with schooner sails.

P. KIRBY

Conceits and provocations

Artists reveal a variety of responses to the contents of a natural-history museum.

Philip Campbell

As you walk through the grounds of Leiden's Naturalis — the national natural-history museum of the Netherlands — you pass compact seventeenth-century buildings, neat plant borders and lawns. You have no idea that under the grass is buried a sculpture by the Los Angeles-based artist Paul McCarthy, valued at around US\$250,000. Or that he himself buried it. Only if you happen to visit the exhibition 'Conversations: Nature and the City', which can be seen at Naturalis until the end of December, will you be told about it and see a video of the event.

The exhibition is the result of an invitation to four visual artists and a composer to respond to the content of Naturalis. This they have done, with installations intended to extend and provoke the visitor's responses to what the museum is collecting and displaying.

McCarthy intends his piece, *Burial*, to reverse the normal process of palaeontologists digging up natural specimens. *Burial*, he explains, "raises a number of questions for natural history museums. While specimens and objects are still buried, do they really exist? Before they're dug up, do they have any value?" To my no-doubt blinkered eye, at least, that question of invisible value is the sort of pseudo-profundity that gives such exercises a bad name.

The exhibition is an experiment inspired by a similar exercise at the Natural History

Museum of Los Angeles County. But the question remains whether it proves its point — that artists' responses and provocations are worth the effort to a museum audience.

Other visual exhibits are conceptually more substantive. Those of Michael McMillen combine a sense of visual play that adds to the cultural value of the specimens. In *The Flying Dutchman*, he places porpoise skeletons in a shoal flying above the visitor's head, with schooner sails attached as a reference to Dutch sea-going. And his amalgamation of an elephant skeleton with a crocodile head surrounded by toy ladders in *Crocodilephant* also provokes the eye. According to the artist: "These natural creations still carry meaning: how we, as a species, reveal our needs, activities, and fears through nature... a carnival for the curious." As the Naturalis organizers suggest, "McMillen's creativity is closer to the way our brains work than we might want to admit."

In *Shank*, Ed Moses displays a large set of stuffed specimens presented on shelves as they would be stored in the museum's vaults, but surrounded by chicken wire. In so doing, he seeks to question the very act of collecting. The piece makes a visual impact and it is good that artists seek to provoke debate among visitors to museum displays. But I found myself wishing that Moses had absorbed his reflections into the subtleties and ambiguities of his art, instead of reducing them to a rather trivial visual concept.

Much more successful for me was John Outterbridge, whose *Sankofa* is a pastel garden of stuffed specimens, bones, plants, crops and artefacts. It has an aesthetic, intriguing appeal. Children demonstrably enjoy it for this reason. And one can readily recognize Outterbridge's description: "Nature in the city, the city in nature... It is a microcosm of our environment, a fertile garden from which we can harvest ideas and reflect on our history and present existence... All of this together signals a living future."

According to Dirk Houtgraaf, the exhibition's curator, attendance of 'Conversations' has been significantly lower than other areas of Naturalis. But I agree with Vanda Vitali, who originated the exhibition in Los Angeles, that such an experiment is destined to be a minority interest. More telling is the behaviour of those visitors who choose to attend. To judge by the attention showed by the visitors I witnessed, the exhibition provides another dimension to the collection's appeal.

What conclusion can be drawn about artists in a museum context? I would suggest that, on this evidence, cumbersome attempts by artists to pose philosophical questions in a visual form tend to smack of conceit, rather than stimulate. But the closer visual artists get to artfulness — to sheer visual creativity — in their response to exhibits, the more likely they are to resonate with the visitors.

Philip Campbell is editor-in-chief of Nature.

NEWS & VIEWS

EVOLUTIONARY BIOLOGY

Ancient genomics is born

David M. Lambert and Craig D. Millar

The reality of a complete Neanderthal genome draws near, as two papers report the sequencing of large amounts of Neanderthal DNA. The results will help to answer some central questions on human evolution.

The study of ancient DNA fascinates everybody. But it has a chequered history, with several high-profile failures — such as false reports of the amplification of DNA sequences from dinosaurs¹ — taking the sheen off undoubted success stories. Among these successes are the identification of human-induced changes in the genes of maize², and the advent of methods for amplifying a single copy of nuclear DNA³. Vocal researchers from the field have argued that it is impossible to sequence the entire genome of an extinct organism, and that the very notion is pure science-fiction because such genomes are typically highly fragmented. How could so many tiny pieces of DNA be sequenced and aligned?

But two papers have now been published that may silence these sceptics. In this issue, Pääbo and colleagues⁴ (Green *et al.*, page 330) describe the sequencing of more than a million base pairs of Neanderthal DNA; for comparison, the human genome is approximately 3.2 billion base pairs long. In collaboration with Pääbo's group, Rubin and colleagues⁵ (Noonan *et al.*) report in *Science* their use of a different method to recover 65,000 base pairs of Neanderthal DNA. These papers are perhaps the most significant contributions published in this field since the discovery of Neanderthals 150 years ago.

The Neanderthals (*Homo neanderthalensis*) were our closest hominid relative. Once widespread throughout Europe and Asia, they became extinct approximately 30,000 years ago. The species began to disappear from the Balkans and central Europe about 40,000 years ago⁶. Modern humans (*Homo sapiens*) are usually blamed, directly or indirectly, for their demise⁷. Nevertheless, it seems likely that humans and Neanderthals learnt about tools and ornaments from each other⁷, and there has been much speculation about possible human–Neanderthal interbreeding, although there is currently no evidence for any such genetic exchange. The results of the new studies^{4,5} further support this conclusion. But sequencing the entire Neanderthal genome would enable us to examine the possibility of Neanderthal interbreeding with humans in much more detail and to explore the genetic basis of functional

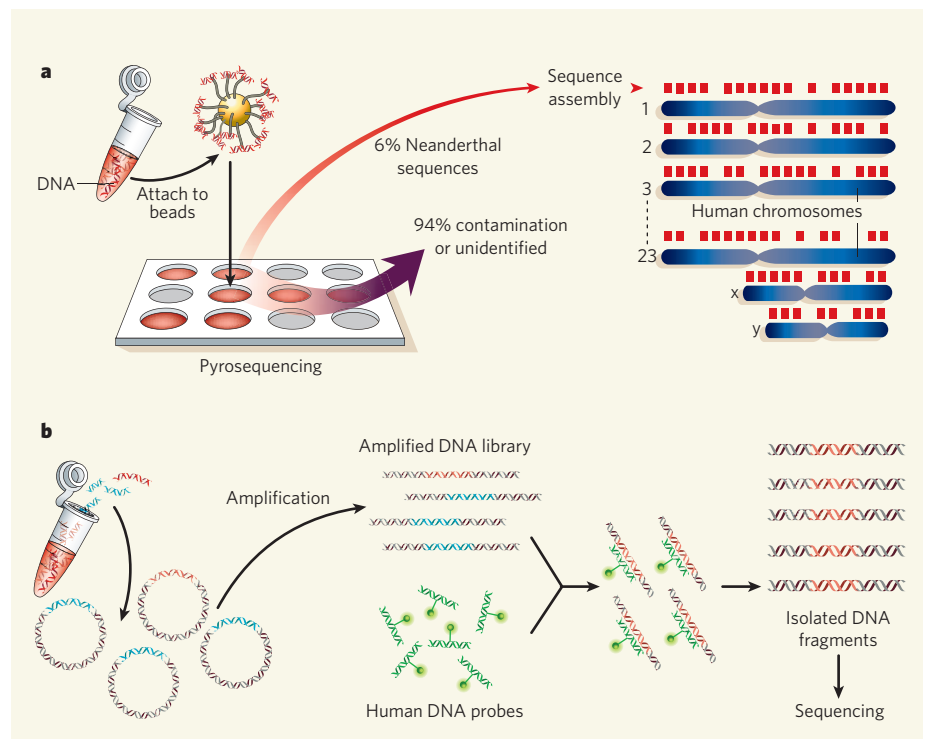


Figure 1 | Sequencing Neanderthal DNA. DNA can be isolated as short fragments from fossilized Neanderthal bones. **a**, Pääbo and colleagues⁴ sequenced more than 1 million base pairs of Neanderthal DNA using an approach known as pyrosequencing. In this method, the isolated DNA fragments are attached to beads that are placed in wells, where the DNA is sequenced directly. About 6% of the sequences obtained were identified by the authors as putative Neanderthal DNA; the remaining sequences were unidentified or represented contamination. The authors localized the Neanderthal sequences (red rectangles) to particular chromosomes by matching them to similar sequences in the human genome. **b**, Rubin and colleagues⁵ used metagenomics to sequence 65,000 base pairs of Neanderthal DNA. In this method, the DNA fragments (red for Neanderthal DNA, blue for contamination) are incorporated into loops of DNA known as plasmids, which are amplified by replicating them in bacteria. The resulting DNA library is then further amplified enzymatically. Because Neanderthal DNA is very similar to DNA from modern humans, the authors used known gene sequences of human DNA to 'trap' complementary strands of the amplified Neanderthal DNA, so isolating them from contaminating sequences. These isolated Neanderthal DNA fragments were then sequenced.

differences between Neanderthals and humans. For example, the Neanderthal genes responsible for cognition and mate recognition would be of great interest.

Pääbo and colleagues⁴ have been able to recover 1 million base pairs of Neanderthal DNA through an advance in DNA sequencing known as pyrosequencing⁸ (Fig. 1a). This technology achieves more than a 100-fold increase

in throughput compared with traditional methods. The latest pyrosequencing machines analyse 25 million DNA bases at a high level of accuracy and in a single four-hour run. Such increased throughput and the ability to directly sequence short fragments of DNA are essential if the entire Neanderthal genome is to be sequenced, as proposed by these authors⁴. In the past, naysayers have insisted that the

reconstruction of ancient genomes is virtually impossible because ancient DNA is typically sheared into short fragments of no more than 100–200 base pairs. But the new pyrosequencing method begins by deliberately breaking genomic DNA into short sequences — so fractured ancient DNA presents no problem.

To reassemble a randomly fractured genome, it is essential to have the complete genome of a closely related organism for comparison. This acts as a kind of scaffold on which isolated DNA sequences can be ‘hung’. Pääbo and colleagues⁴ used the human genome, mapping Neanderthal DNA sequences to particular regions on the human chromosomes. Studies of the extinct mammoth⁹ used the elephant genome in a similar way. The sequenced genomes from mice, fruitflies and chickens are likely to act as DNA scaffolds for a range of extinct organisms in the future.

With more than 1 million base pairs analysed, the authors⁴ compared their Neanderthal DNA sequence with the chimpanzee and human genomes. Assuming that humans and chimpanzees diverged 6.5 million years ago, and using a combination of likely population sizes and divergence times, Pääbo and colleagues⁴ estimate that human and Neanderthal DNA diverged between 465,000 and 569,000 years ago, with a best estimate of about 516,000 years ago. Interestingly, the authors also show that Neanderthals were derived from a very small ancestral population of about 3,000 individuals. This is similar to the estimated ancestral population size of humans — that is, the number of individuals needed to produce the amount of observed genetic diversity within the current population.

Rubin and colleagues⁵ used a ‘metagenomic’ approach to sequence the Neanderthal genome (Fig. 1b). Metagenomic techniques were originally developed to recover genomes from environmental samples, enabling the study of organisms that are not easily grown in the laboratory. The authors have adapted this method to recover large tracts of Neanderthal DNA. Their results agree broadly with those of Pääbo and colleagues⁴ — for example, Rubin and colleagues estimate the divergence time for humans and Neanderthals to be between 120,000 and 670,000 years ago, with their best estimate being 370,000 years ago.

The metagenomic approach⁷ recovered only small amounts of Neanderthal DNA compared with the pyrosequencing⁴ method. But metagenomics does offer some advantages, as it enables Neanderthal sequences that are similar to human genes to be targeted. Crucially, Rubin and colleagues⁵ recovered 29 out of 35 genes that were targeted in this way. So although pyrosequencing recovers larger amounts of ancient sequence, metagenomics allows the focused recovery of specific regions of interest. Which of the two methods will be the more appropriate for future applications will depend on the question being addressed.

Although these studies^{4,5} will not immediately

answer many of the questions about the biological differences between Neanderthals and humans, they foreshadow an exciting development — the recovery of the complete Neanderthal genome. Pääbo and colleagues⁴ promise this within two years, and propose to sequence the Neanderthal genome six times over. This will allow any rare mistakes in sequencing to be identified using the consensus from independent reads. Multiple sequencing of genomes is standard practice — for example, the human genome was sequenced, on average, ten times over. But a big difficulty for Pääbo’s team is the cost. Each pyrosequencing run is very expensive, and the authors estimate that 6,000 runs are required. However, the authors allude⁴ to improvements in the technique that will reduce the number of runs tenfold.

The full Neanderthal genome will provide a precise assessment of differences between us and our closest relative. It may also help to resolve a debate that started more than 20 years ago¹⁰, when studies of ancient DNA began: in evolution, how important are mutations in genes that result in structural and physiological changes, compared with mutations that affect

the regulation of those genes? More fundamentally, these combined studies show that, when predicting the limits of science, one should never say never.

David M. Lambert is at the Allan Wilson Centre for Molecular Ecology and Evolution, Institute of Molecular BioSciences, Massey University, Private Bag 102 904 NSMC, Auckland, New Zealand.

e-mail: d.m.lambert@massey.ac.nz

Craig D. Millar is at the Allan Wilson Centre for Molecular Ecology and Evolution, and the School of Biological Sciences, University of Auckland, Private Bag 920109, Auckland, New Zealand.

e-mail: cd.millar@auckland.ac.nz

1. Woodward, S. R., Weyand, N. J. & Bunnell, M. *Science* **266**, 1229–1232 (1994).
2. Jaenicke-Despres, V. et al. *Science* **302**, 1206–1208 (2003).
3. Huynen, L., Millar, C. D., Scofield, P. R. & Lambert, D. M. *Nature* **425**, 175–178 (2003).
4. Green, R. E. et al. *Nature* **444**, 330–336 (2006).
5. Noonan, J. P. et al. *Science* **314**, 1113–1118 (2006).
6. Pettiitt, P. B. *Oxf. J. Archaeol.* **18**, 217–240 (1999).
7. Jankovic, I. *Coll. Anthropol.* **28** Suppl. 2, 379–401 (2004).
8. Margulies, M. et al. *Nature* **437**, 376–380 (2006).
9. Poinar, H. N. et al. *Science* **311**, 392–394 (2006).
10. Higuchi, R., Bowman, B., Freiberger, M., Ryder, O. A. & Wilson, A. C. *Nature* **312**, 282–284 (1984).

SEISMOLOGY

Greatness thrust upon them

James F. Dolan

The latest research seems to imply that all earthquakes are born equal. But combining that insight with earlier, seemingly contradictory, work could help us to tell which tremors grow to become more equal than others.

In an analysis on page 358 of this issue¹, Steven Wesnousky provides strong evidence that the ultimate size of a seismic rupture is largely controlled by the structure of the underlying fault, and therefore that big earthquakes do not differ from small earthquakes in their beginnings. These data might seem to conflict with earlier observations implying that the size of an earthquake is determined by the dynamics of rupture onset. In fact, both conclusions could be true, and combining these two data sets in future analyses of seismic hazards might result in a better prediction of the eventual size of an earthquake before the shaking stops.

Many Earth scientists have long suspected that the limit of an earthquake rupture, and therefore the magnitude of an event, is largely controlled by the structure of the fault zone and variations in stress along the fault itself^{2–4}. In this view, all earthquakes begin in the same way, and continue to propagate until they reach a barrier, in the form of a structural complexity or a part of the fault on which the stresses are sufficiently low to stop rupture. But the alternative standpoint — that large earthquakes are in some way born differently from their smaller

brethren^{5,6} — is attractive because it holds the promise of determining the eventual size of an earthquake during the first few seconds of the rupture. That could provide the basis for an effective early-warning system.

Wesnousky’s contribution¹ to this debate is to examine the end-points of 22 surface-rupturing earthquakes of moderate to large magnitude ($M = 6.1–7.9$) from the past 150 years. He found that approximately two-thirds of all ruptures terminated at either a structural ‘step’ in the fault or at an intersection with a second fault. Moreover, he found that no ruptures, regardless of their magnitude, were able to propagate through structural steps wider than 3–4 kilometres. Collectively, these observations demonstrate the fundamental role played by the structure of the fault in stopping a seismic rupture, and imply forcefully that an earthquake cannot ‘know’ how large it is destined to be until forced to stop at a structural or stress barrier.

These observations are in apparent conflict with results showing that the predominant frequency of the waves of seismic energy radiated during the first few seconds of a rupture

Figure 1 | At fault. The 400-km-long rupture along the Kunlun fault on the Tibetan plateau that occurred in 2001 was the longest of those studied by Wesnousky¹. According to his analysis, the ultimate size of an earthquake is primarily determined by the local fault structure.



HOU DEQIANG/XINHUA

differs for large and small earthquakes. Olsen and Allen⁶, for example, examined the frequency content of the early stages of 71 seismic ruptures from $M = 3$ to $M = 8.3$. The predominant frequency of these earthquakes decreased with increasing magnitude; in other words, small earthquakes released a higher proportion of their energy as higher-frequency waves than did large earthquakes. That seemed to indicate that the final magnitude of an earthquake is determined before the end of the event, and so is at least partially controlled by the rupture initiation process.

An explanation for the seeming contradiction of these conclusions^{1,6} could lie in the detailed structure of the fault planes themselves. Earlier research, much of it by Wesnousky and colleagues^{7,8}, showed that the more a fault slips over its lifetime, the smoother the slip surface becomes. The structure of faults with small displacements is therefore likely to be much more complex than that of faults that have accommodated tens to hundreds of kilometres of displacement.

It is not surprising that earthquakes nucleated on smooth faults generate less energy at high frequencies right from the start than do earthquakes nucleated on rougher, less-developed faults. Each small-scale structural feature of a fault acts as a separate source of high-frequency energy. Olsen and Allen's correlation⁶ might thus simply reflect the fact that small earthquakes can occur on small faults that have not accommodated significant displacements over their lifetimes. Many of these earthquakes may occur in the zone of damaged rock surrounding large faults.

Large earthquakes, by contrast, necessarily

occur on large faults — the largest earthquake examined in Wesnousky's study¹, an event of $M = 7.8$ in 2001, ruptured the Kunlun fault on the Tibetan plateau for a distance of more than 400 kilometres (Fig. 1). Such large faults are much more likely to have accommodated significant displacements, and are likely to be much smoother than the innumerable small faults that generate the earthquakes of $M < 5.5$ that are ubiquitous features of Earth's crust.

Interestingly, although the relationship between magnitude and wave frequency noted by Olsen and Allen⁶ is quite pronounced for smaller magnitudes between 3 and 5.5, there is no obvious variation in the predominant frequency of energy radiated in the first few seconds of $M > 6$ earthquakes. The radiation of energy at relatively low frequencies for these larger earthquakes could tell us which of the ruptures are starting on large, smooth faults, and therefore might be capable of generating large earthquakes — if the structural and stress conditions of the fault are favourable.

Wesnousky's data¹, on the other hand, show that the ultimate size of the earthquake is largely controlled by the structure of the fault zone, rather than by the dynamics of the rupture nucleation. They thus validate the continued use of mapped structural complexities (steps and bends in the fault surface, intersections with other faults) in forward modelling of the potential limits of future ruptures. Such data are the basis for many seismic-hazard models in use today⁹.

Thus, combining these two sets of observations^{1,6} in automated form could result in more accurate assessments of likely earth-



50 YEARS AGO

It is not without significance that the opening by the Queen of Britain's first nuclear power station has almost coincided with the approval by a conference of eighty-two nations of the draft statute for the International Atomic Energy Agency... The urgent need is for some arrangement whereby the technical skill of Britain and those other nations within reasonable distance of producing nuclear power on a commercial scale can be made available to those less fortunate nations; but it still remains to be seen how widely the new statute will be ratified by governments, and in particular how ready the smaller and undeveloped countries will be to accept the provisions of Article 12 of the statute... The safeguards imposed by that article and the limitations of national sovereignty implied in it have become the more essential and reasonable; yet...in many parts of the world where nuclear power offers the greatest possibilities of benefit to human welfare, the growth of nationalism has made those countries more impatient and suspicious of restraint.

From *Nature* 17 November 1956.

100 YEARS AGO

A statement has recently obtained currency that the French people themselves, after a hundred years' use of the metric system, cannot claim that it has been adopted throughout France, and a free translation of a circular issued to chambers of commerce in France by the French Minister of Commerce has been employed to support the statement... [The Minister] makes it clear that the circular is directed only against the use of *old names* in certain trades, and that the English translation misinterprets its meaning and conveys a wholly wrong impression. It is satisfactory to find, in view of such endeavours to retard the acceptance of the metric system by this country, that it has recently been adopted in the works of Messrs. Joseph Crosfield and Sons, Ltd., and steadily grows in popularity.

From *Nature* 15 November 1906.

50 & 100 YEARS AGO

quake size before the shaking stops. The frequency content would identify an earthquake that could grow large, and the structure of the fault would identify the likely end-point of the rupture. Knowing in real time the difference between a moderate magnitude-6 earthquake and a truly destructive magnitude-8 one on the San Andreas fault near Los Angeles, for example, could be important indeed.

James F. Dolan is in the Department of Earth Sciences, University of Southern California, Los Angeles, California 90089, USA.

e-mail: dolan@usc.edu

1. Wesnousky, S. G. *Nature* **444**, 358–360 (2006).
2. Segall, P. & Pollard, D. D. *J. Geophys. Res.* **85**, 4337–4350 (1980).
3. Sibson, R. H. *Nature* **316**, 248–251 (1985).
4. Harris, R. A. & Day, S. M. *Geophys. Res. Lett.* **26**, 2089–2092 (1999).
5. Ellsworth, W. L. & Beroza, G. C. *Science* **268**, 851–855 (1995).
6. Olsen, E. L. & Allen, R. M. *Nature* **438**, 212–215 (2005).
7. Wesnousky, S. *Nature* **335**, 340–343 (1988).
8. Stirling, M. W., Wesnousky, S. G. & Shimazaki, K. *Geophys. J. Int.* **124**, 833–868 (1996).
9. Frankel, A. D. et al. *US Geol. Surv. Open-File Rep.* 02-420 <http://pubs.usgs.gov/of/2002/ofr-02-420> (2002).

sensitive to these perturbations (intermolecular paramagnetic relaxation enhancement), they show that a variety of transient, non-specific encounter complexes of the two proteins are present under equilibrium conditions. Neither the final, specific complex structure (previously solved by NMR) nor any other single alternative conformation alone can account fully for their data.

This paramagnetic approach has already proved helpful for detecting rare equilibrium intermediates in protein–DNA interactions⁴. But Tang and colleagues² go farther, and use computation-based structural models to build an ensemble of conformations that satisfy the experimental data. They thus provide structural evidence for the rare encounter complexes, and suggest that electrostatic interactions have a major role in forming these intermediates (Fig. 1; see also Fig. 3a of the paper² on page 385).

Interactions between proteins are highly prevalent in biology, and the sorts of transient interaction revealed by Tang *et al.* are not captured in the crystals that are the basis of many structural studies, so this dynamic approach will be useful in many other cases. Indeed, populations of alternative encounter complexes will probably be even more evident in highly transient protein–protein interactions⁵, such as those involved in the inter-protein electron-transfer processes that occur in cellular metabolism. For instance, in the interaction between the Cu_A domain of the *ba*₃ oxidase and its interaction partner cytochrome *c*₅₅₂, a single complex cannot fully account for the restraints on the amino acids observed in NMR-based experiments, leaving open the possibility that an ensemble of conformations coexist in a weakly complexed state⁶.

Many fundamental questions remain about the structure and energetics of these encounter complexes. The role of long-range electrostatic forces in bringing molecules together before they collide has been studied from both experimental and theoretical viewpoints^{7,8}, and it

CELL BIOLOGY

Brief encounters bolster contacts

Tom L. Blundell and Juan Fernández-Recio

Molecules often work together in complexes to carry out their functions in the cell. But how do they get together in such a dynamic environment? A structural study follows proteins as they meet their partners.

Living cells, particularly during growth and proliferation, need regulatory processes of great sensitivity and high specificity. To achieve this, signal-to-noise ratios must be high when information is received and transmitted between the cell surface, the cytoplasm and the nucleus. Just like electrical and engineering control systems, living cells have complex signalling pathways that are moderated by feedback mechanisms. It is becoming increasingly clear that most switches, transducers and adaptors in living systems are created by the assembly and disassembly of multi-component complexes of proteins, nucleic acids and other molecules¹. On page 383 of this issue, Tang *et al.*² follow the formation of such complexes, and provide structural evidence for the transient interactions necessary to build them*.

How do the molecular assemblies in cells achieve the required sensitivity and specificity? Efficient signal transduction must maintain fidelity and decrease noise while amplifying the signal. So the solution cannot be explained in terms of tightly bound, enduring molecular complexes, because the signals could not then be turned off. Rather, it seems to lie in first assembling weak binary complexes, and then using cooperative interactions to produce multi-component complexes in which the weak interactions are replaced by much stronger and more specific interactions³.

Although weak, nonspecific, transient complexes could give rise to a noisy system, such 'encounter complexes' might be exploited so that interaction partners do not have to be found afresh in the busy milieu of the cell, thus increasing the rate of formation of specific

binary and higher-order complexes. Essentially, the partners bump into one another and are held loosely, allowing them time to become reorientated and repositioned on the surface or to adjust their shape to fit together more tightly. Recent studies are beginning to describe the dynamics of the assembly processes and to show that nonspecific, transient collisions play an important role in macromolecular associations.

Tang *et al.*² study a bacterial signalling system called the phosphotransferase system, and describe evidence for rare encounter complexes involving the amino-terminal domain of enzyme I and the phosphocarrier protein HPr. The authors mutated one of the proteins at specific sites, introducing a paramagnetic atom that causes local perturbations in the magnetic field. Using a technique that is

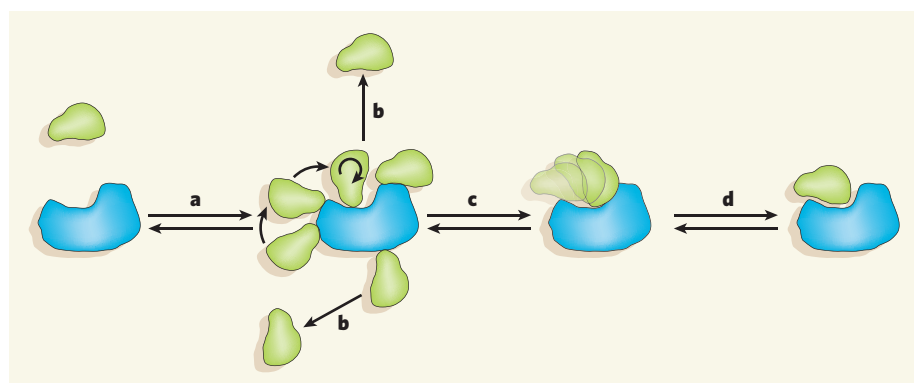


Figure 1 | Protein–protein interactions. Equilibrium steps in a possible mechanism for protein–protein association, compatible with data from Tang *et al.*². **a**, Formation of transient encounter complexes by nonspecific collisions, guided mostly by electrostatic interactions. **b**, Many encounter complexes separate rapidly. **c**, Some productive encounter complexes reorientate and come closer to the final, specific orientation, guided mostly by desolvation, as water molecules move away from the protein surfaces. **d**, Formation of the specific complex, with final fitting of interacting surfaces.

*This article and the paper concerned² were published online on 15 October 2006.

is clearly a major driving force in the formation of the rare encounter complexes described by Tang and colleagues. Less studied, however, is the role of short-range desolvation effects — where the protein–protein interactions force solvent molecules away from the proteins — during and after collisions. It has been suggested that hydrophobicity is involved in reorientating the molecules to form the final, productive complex⁵. Computational simulations also show that desolvation energy plays a part in orientating the encounter complexes' interacting subunits around the final, specific complex state⁹.

Tang *et al.*² describe structural features of the encounter complexes (assuming that they interact as rigid bodies), and their results are consistent with the general idea of a funnel-shaped binding-energy well that narrows as the two proteins approach one another. This implies that there are many possible routes for arriving at the final complex at the bottom of the energy well, and that these are determined by transient interactions between the partners in the encounter complexes, with the pathways converging as they get lower in energy and closer to the final complex. Indeed, rigid-body docking calculations based on optimization of the binding energy of the interacting molecules^{9–11} already describe a pool of alternative encounter complexes on the way to forming the functional complex. Whether these ensembles of orientations reflect the true binding-energy landscape will depend on the accuracy of the energy description of these computer models and the efficiency of the sampling method, an area of current debate. Molecular-dynamics simulations show that some encounter complexes could be sufficiently long-lived for their side chains to acquire a variety of conformational states, some of which are similar to those in the final, functional complex¹². But it remains to be seen how many of the minor species are true productive encounter complexes, and which are the preferred paths to the specific binding mode of the final complex.

It is now apparent that rare encounter complexes might control not only the kinetics of the assembly process, but also the way the complex is put together and hence its cooperativity. Furthermore, the population of non-specific complexes can be restricted by the order in which the different subunits are assembled. Greater understanding of the route to longer-term relationships between molecules will no doubt emerge from integrating a wide variety of experimental data with theoretically sound computer modelling^{13,14} of their brief encounters. ■

Tom L. Blundell is in the Department of Biochemistry, University of Cambridge, Tennis Court Road, Cambridge CB2 1GA, UK. Juan Fernández-Recio is at the Institute of Biomedical Research, Barcelona Science Park, Barcelona 08028, Spain.
e-mail: tom@cryst.bioc.cam.ac.uk

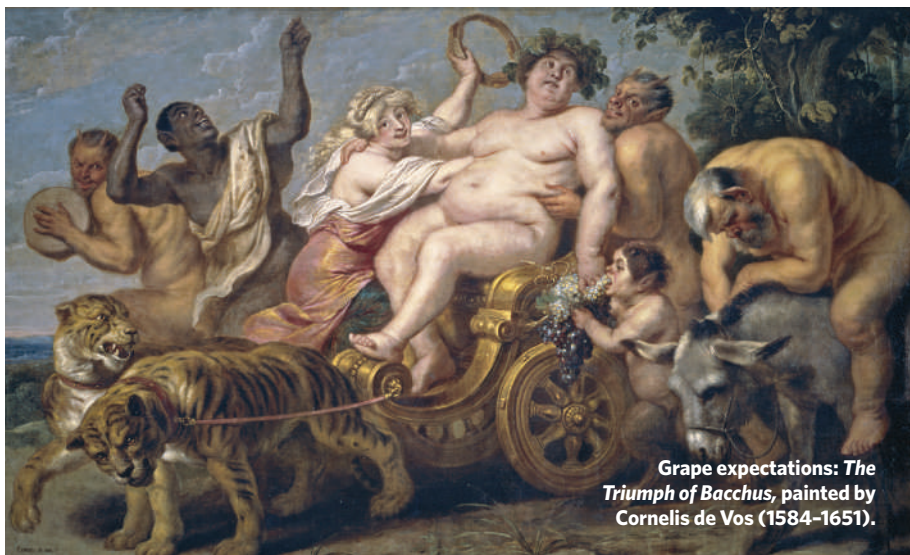
1. Gavin, A. C. *et al.* *Nature* **415**, 141–147 (2002).
2. Tang, C., Iwahara, J. & Clore, G. M. *Nature* **444**, 383–386 (2006).
3. Harmer, N. J. *et al.* *Biophys. Chem.* **100**, 545–553 (2002).
4. Iwahara, J. & Clore, G. M. *Nature* **440**, 1227–1230 (2006).
5. Crowley, P. B. & Ubbink, M. *Acc. Chem. Res.* **36**, 723–730 (2003).
6. Muresanu, L. *et al.* *J. Biol. Chem.* **281**, 14503–14513 (2006).
7. Schreiber, G. & Fersht, A. R. *Nature Struct. Biol.* **3**, 427–431 (1996).
8. Gabbouline, R. R. & Wade, R. C. *Curr. Opin. Struct. Biol.* **12**, 204–213 (2002).
9. Fernández-Recio, J., Totrov, M. & Abagyan, R. *J. Mol. Biol.* **335**, 843–865 (2004).
10. Gray, J. J. *et al.* *Proteins* **52**, 118–122 (2003).
11. Camacho, C. J. & Vajda, S. *Proc. Natl Acad. Sci. USA* **98**, 10636–10641 (2001).
12. Rajamani, D., Thiel, S., Vajda, S. & Camacho, C. J. *Proc. Natl Acad. Sci. USA* **101**, 11287–11292 (2004).
13. de Bakker, P. I., Furnham, N., Blundell, T. L. & DePristo, M. A. *Curr. Opin. Struct. Biol.* **16**, 160–165 (2006).
14. Furnham, N., Blundell, T. L., DePristo, M. A. & Terwilliger, T. C. *Nature Struct. Mol. Biol.* **13**, 184–185 (2006).

MEDICINE

Grapes versus gluttony

Matt Kaeberlein and Peter S. Rabinovitch

A compound found in red grapes called resveratrol improves the health and lifespan of mice on a high-calorie diet. This is potentially good news for overweight humans. Does it bode well for the rest of us too?



Grape expectations: The Triumph of Bacchus, painted by Cornelis de Vos (1584–1651).

Bacchus (Dionysus to the Greeks) has been long out of style, but may be granting new favours — particularly if you long to be one of those people who can seemingly eat whatever they want, whenever they want, without having to worry about the consequences. A paper by Baur *et al.*¹ on page 337 of this issue suggests that guilt-free gluttony might not be a fantasy*.

In this report, mice fed a diet akin to coconut cream pie for every meal showed a striking increase in survival and health when their chow was supplemented with resveratrol, a polyphenolic compound found in red grapes or wine. Compared with animals fed a more standard diet, mice fed the high-calorie (60% from fat) diet without resveratrol had a shorter lifespan. They also showed many of the problems that plague humans who overindulge at the dinner table, including obesity, insulin resistance and heart disease. Baur *et al.* found

*This article and the paper concerned¹ were published online on 1 November 2006.

that although resveratrol did not prevent obesity, it did prevent obesity-associated disease, at least in one strain of mouse, and conferred a nearly normal lifespan on these mice.

With the present epidemic of obesity in some Western societies, this could be very good news. But might resveratrol improve health or lifespan beyond that achieved with a healthy diet? The link between diet and longevity has been known to gerontologists since the discovery in the 1930s that reduced caloric intake can increase the lifespan of rodents by up to 50%. Dietary restriction has since been observed to have a similar effect on longevity in many different organisms, including yeast, worms, flies, spiders and fish. Importantly, dietary restriction not only increases lifespan, but it also delays the onset of nearly all age-associated diseases. For this reason, most gerontologists believe that dietary restriction affects the intrinsic ageing process at a fundamental level. The genetic pathways influencing this phenomenon are currently

a hot topic of research and debate.

Like dietary restriction, resveratrol has long been known to have interesting properties. During the 1990s it was extensively studied as a potential link between improvements in a variety of health indicators and moderate consumption of red wine². The antioxidant properties of resveratrol, in particular, have been suggested to account for many of its beneficial properties, including putative cardio-protective and anticancer activities, as well as providing protection against liver failure. Here it is noteworthy that Baur *et al.*¹ show that resveratrol has a profound ability to prevent liver damage associated with the high-fat diet.

Resveratrol became of particular interest to gerontologists with the report³ that it can increase lifespan in yeast by activating particular enzymes (protein deacetylases) of the Sir2 family of proteins (sirtuins). Sirtuins are evolutionarily conserved mediators of longevity that might also play a role in lifespan extension through dietary restriction⁴. Although the results from the initial study of resveratrol in yeast remain controversial⁵, subsequent work has suggested that resveratrol has modest effects on lifespan in both worms and flies⁶, and a more substantial effect on lifespan in a short-lived fish⁷. Based on these findings, it has been proposed that resveratrol increases lifespan in several different organisms by a mechanism similar to dietary restriction⁸.

Baur *et al.*¹ favour the view that many (perhaps all) of the beneficial properties of resveratrol are the result of increased sirtuin activity, and various studies have supported the idea that sirtuins underlie the effects attributed to resveratrol *in vivo*⁸. However, there is a surprising lack of biochemical evidence that resveratrol directly increases sirtuin-mediated deacetylation of biologically relevant substrates, and some evidence that it may not^{5,9}. Resveratrol is also known to interact with numerous proteins and pathways, including mitochondrial ATP synthase and complex III, fatty-acid synthase, protein kinase C, p53, MEK1, TNF- α and NF- κ B, all of which are candidates for mediating its *in vivo* effects. In particular, activation of AMP kinase by resveratrol protects against atherosclerosis and liver damage in diabetic mice¹⁰, suggesting a likely mechanism for the observations reported by Baur and colleagues.

Given the available data, it is difficult to predict the answers to a few key questions. Will resveratrol have an effect on health and longevity in mice fed a standard diet, rather than a high-calorie diet? Will it be effective in mice with genetic backgrounds other than the inbred strain used in the current report? Will it be effective in humans? Studies addressing these questions are under way: the answers will go some way towards determining whether or not resveratrol is a bona fide dietary-restriction mimetic.

Many people will wonder whether they should start supplementing their diets with

resveratrol. After all, it is generally regarded as safe, and can be purchased over the Internet with promises of improved health and longevity. Our advice is to exercise caution. The safety of resveratrol at the high doses in humans comparable to those used by Baur *et al.*¹ is unknown, especially over the course of years or even decades, when relatively modest side effects could have dramatic consequences. A logical next step would be to initiate controlled studies to find out whether resveratrol can safely reduce the ill-effects associated with diabetes or obesity in humans.

In the most optimistic assessment, a true mimetic of dietary restriction could be effective against many age-associated human diseases, including heart disease, diabetes, cancer and neurological disorders such as Alzheimer's disease. Even if resveratrol doesn't make the grade, it is not the last hope of gerontologists, or necessarily even the best. Studies of several other compounds are under way in multicentre studies of mouse ageing sponsored by the National Institute on Aging¹¹. These include potent antioxidants and compounds targeting other pathways thought to influence lifespan extension through dietary restriction.

For now, we counsel patience. Just sit back and relax with a glass of red wine — which, alas, has only 0.3% of the relative resveratrol dose given to the gluttonous mice (note also that increasing the dose via wine will not be healthy). But if you must have a Big Mac, fries and apple pie, we may soon know if you should supersize that resveratrol shake. ■

Matt Kaeberlein and Peter S. Rabinovitch are in the Department of Pathology, University of Washington, Seattle, Washington 98195, USA. e-mails: kaeber@u.washington.edu; petersr@u.washington.edu

1. Baur, J. A. *et al.* *Nature* **444**, 337–342 (2006).
2. Soleas, G. J., Diamandis, E. P. & Goldberg, D. M. *Clin. Biochem.* **30**, 91–113 (1997).
3. Howitz, K. T. *et al.* *Nature* **425**, 191–196 (2003).
4. Longo, V. D. & Kennedy, B. K. *Cell* **126**, 257–268 (2006).
5. Kaeberlein, M. *et al.* *J. Biol. Chem.* **280**, 17038–17045 (2005).
6. Wood, J. G. *et al.* *Nature* **430**, 686–689 (2004).
7. Valenzano, D. R. *et al.* *Curr. Biol.* **16**, 296–300 (2006).
8. Baur, J. A. & Sinclair, D. A. *Nature Rev. Drug Discov.* **5**, 493–506 (2006).
9. Borra, M. T., Smith, B. C. & Denu, J. M. *J. Biol. Chem.* **280**, 17187–17195 (2005).
10. Zang, M. *et al.* *Diabetes* **55**, 2180–2191 (2006).
11. www.nia.nih.gov/ResearchInformation/ScientificResources/InterventionsTestingProgram.htm

FLUID DYNAMICS

Spinning discs in the lab

Steven A. Balbus

What causes gas to be drawn in towards black holes, rather than remain in a stable orbit as planets do around the Sun? A laboratory result indicates that something more than just hydrodynamics must be at work.

On page 343 of this issue, Ji *et al.*¹ describe a meticulous experiment in which they confined water between two independently turning cylinders. Through artful experimental design, the authors were able to reduce viscous effects in the resulting 'Couette' flow to a level of one part in two million. They chose the velocities of the cylinders so that they would mimic — and so compel the confined fluid to mimic — so-called keplerian rotation, which is typical of astrophysical disks around black holes. Here, velocity is inversely proportional to the square-root of the distance from the centre of the rotation.

The result was that nothing happened at all: the fluid continued to rotate stably. But why exactly do astrophysicists and fluid dynamacists find this apparently harmless result so surprising?

In the early 1970s, astrophysicists were struggling with the exciting and controversial question of whether black holes — objects whose gravity is so great that nothing, not even light, can escape once captured — were real². Only slightly earlier, a series of compact sources of X-ray radiation had been discovered. One

model held that this radiation originated from gas disks surrounding black holes in binary systems of close stars³ and galactic nuclei⁴. This gas would dissipate its energy as heat, and ultimately X-ray radiation. Its angular momentum would be transported outward, and the gas would spiral inward towards the hole. By understanding these 'accretion disks', astronomers hoped that they would, in one fell swoop, both explain the mysterious X-ray sources and prove that black holes exist.

The existence of black holes and their accretion disks is now widely accepted by both theorists and observers. But understanding the dynamics of accretion disks, in black holes and in other types of system, has turned out to be an extremely knotty problem. Why do disks accrete at all? Why does gas in motion around a massive centre not remain in a stable, planet-like orbit?

The problem is that energy dissipation and angular-momentum transport are properties of a viscous fluid, and the viscosity of the disk gas is far too small to account for the angular-momentum loss that leads to accretion. This problem might be solved if a keplerian gas

were turbulent. A turbulent fluid can have a small particulate viscosity, yet behave in some ways as though it were viscous: correlations between different components of the fluid's velocity fluctuations cause both substantial energy dissipation and enhanced momentum transport.

The importance of viscosity in fluid flow is measured by a dimensionless quantity known as the Reynolds number; a high Reynolds number indicates a small viscosity. Shear flows (those with regions moving at different velocities) of high Reynolds number are generally exquisitely unstable to any small perturbation⁵. But astrophysical disks both shear and rotate; the rotation introduces Coriolis forces, which can sometimes stabilize an otherwise turbulent fluid.

Starting in the late nineteenth century, the British physicist Lord Rayleigh investigated fluids similar to the gas in an astrophysical disk that rotate differentially (at different speeds according to radial distance) and have a very small viscosity. He found that such fluids would become unstable (and possibly turbulent) only where the specific angular momentum of the flow decreases as one moves to greater radii. In a keplerian disk, however, the specific angular momentum increases with radius. According to the Rayleigh criterion, such a flow is stable.

Because the Rayleigh criterion applies only to infinitesimal disturbances, not those of finite amplitude, and only to perturbations that are symmetrical about the axis of rotation, the formal stability of keplerian disks did not immediately embarrass the theorists. Indeed, although Couette experiments⁶ persistently failed to reveal instability for keplerian profiles, it was widely believed that differentially rotating flows would ultimately prove to be unstable; it was simply a matter of reaching large enough Reynolds numbers in the laboratory⁷.

For many, this 'faith-based' approach to turbulent hydrodynamical accretion was less than satisfying. The discovery⁸ in 1991 that even very weak magnetic fields profoundly alter the stability of rotating gases improved matters. A rotating magnetized gas becomes unstable when its angular velocity decreases as one moves away from the centre, a condition nearly universally satisfied in astrophysical disks. Computer simulations have since shown that this 'magnetorotational instability' leads to precisely the sort of turbulence that disk theorists are seeking⁹.

Not all disks are ionized throughout to an extent that would allow magnetic fields the necessary influence to seed instability. Only a very small electron component is needed to couple the field and the gas, but accretion disks associated with the formation of stars, known as protostellar disks, seem to contain an extended region that is cold, dense and dusty. In such a region, free electrons are suppressed, and gas and field are decoupled. For this reason and others, many disk theorists continued to

believe that a fluid of large Reynolds number undergoing keplerian shear would ultimately prove to be turbulent. In 2001, a laboratory Couette-flow experiment at large Reynolds number seemed to find just that¹⁰.

One of the great difficulties in working with Couette flows is that the rotating cylinders have end-caps that rotate uniformly, whereas in the adjacent fluid the rotation is a function of distance from the rotation axis. This mismatch creates a boundary layer near the end-cap and a secondary axial velocity flow known as an Ekman circulation in addition to the purely rotational flow that one wishes to study. This effect can be controlled to some extent by making the cylinders very long, distancing the end-caps from the bulk of the flow.

Ji and colleagues, however, wish to use their apparatus¹ in investigations of the magnetorotational instability. These require a very uniform magnetic field that would be difficult to maintain over a long cylinder. So the authors overcame the Ekman circulation problem by splitting the end-caps into two parts that rotate at different speeds. This was sufficient to control the Ekman circulation and to attain an exceptionally high Reynolds number of around 2×10^6 .

The authors' finding that, at such a high Reynolds number, the rotation profile of a keplerian fluid is as stable as the rotation of a solid body should lay to rest the notion that any rotating flow (in general) or an accretion disk (in particular) is nonlinearly unstable

simply if its viscosity is sufficiently small. The implication, for which there is now growing astrophysical evidence¹¹, would seem to be that accretion is magnetically, not hydrodynamically, driven.

Ji and colleagues' primary goal for their apparatus was to find evidence for the magnetorotational instability in liquid gallium, and the results of these experiments are yet to come. The researchers might thus be poised for a double coup: tolling the death knell for hydrodynamical shear turbulence in accretion disks, and capturing magnetic instability in a cylindrical flow for the first time. ■

Steven A. Balbus is at the Laboratoire de Radioastronomie, Ecole Normale Supérieure, 24 rue Lhomond, 75231 Paris cedex 05, France. e-mail: steven.balbus@lra.ens.fr

1. Ji, H., Burin, M., Scharfman, E. & Goodman, J. *Nature* **444**, 343–346 (2006).
2. Novikov, I. D. & Thorne, K. S. in *Black Holes: Les Astres Occlus* (ed. De Witt, C.) 343 (Gordon & Breach, New York, 1973).
3. Prendergast, K. H. & Burbidge, G. R. *Astrophys. J.* **151**, L83–L88 (1968).
4. Lynden-Bell, D. *Nature* **223**, 690–694 (1969).
5. Drazin, P. G. & Reid, W. H. *Hydrodynamic Stability* (Cambridge Univ. Press, 1981).
6. Coles, D. J. *Fluid Mech.* **21**, 385 (1965).
7. Landau, L. D. & Lifshitz, E. M. *Fluid Mechanics* (Pergamon, Oxford, 1959).
8. Balbus, S. A. & Hawley, J. F. *Astrophys. J.* **376**, 214–222 (1991).
9. Hawley, J. F., Gammie, C. F. & Balbus, S. A. *Astrophys. J.* **440**, 742–763 (1995).
10. Richard, D. Thesis, Univ. Paris 7 (2001).
11. Miller, J. M. *et al. Nature* **441**, 953–955 (2006).

ENVIRONMENTAL CHEMISTRY

Browning the waters

Nigel Roulet and Tim R. Moore

Levels of dissolved organic carbon in British streams and lakes have risen over the past two decades. It might be a downstream effect of decreased acid rain — but isolating single factors is notoriously difficult.

It may be small, but dissolved organic carbon (DOC) — operationally defined as organic compounds in water that can pass through a 0.45- μm filter — is of great interest. The export of DOC from land to aquatic ecosystems and the oceans is a significant mover of carbon through continents and local landscapes (Fig. 1, overleaf). In some ecosystems, loss of DOC can be as great as long-term accumulation of carbon in the form of dead organic material, for example as peat in peatlands. Then there is its role in the metabolism of lake ecosystems and the protection of aquatic organisms: DOC-rich waters are light brown, and so absorb ultraviolet radiation. On the other hand, DOC-rich water can produce carcinogens when chlorinated. Last — but for the bibulous not least — DOC content is even believed to help determine differences

in the flavours of malt whiskies.

This is the context in which Evans *et al.*, writing in *Global Change Biology*¹, supply evidence that median DOC concentrations in eight streams and ten lakes in the United Kingdom have almost doubled from 1988 to 2003. It is important to distinguish between DOC export and concentration here: whereas increased export can result from an increase in runoff with no change in concentration, increased concentrations can occur with no change in hydrology, but with altered production or retention of DOC in the landscape. Increased export of DOC through streams and rivers in the Northern Hemisphere has been observed before, specifically in western Siberia as a result of increased release of DOC from peatlands². But decreased DOC export has also been observed, in the Yukon river system of

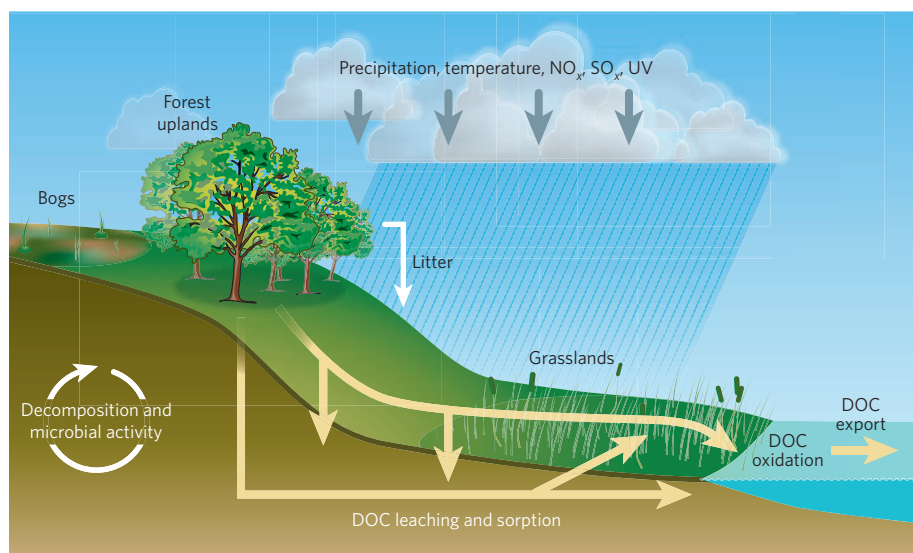


Figure 1 | Passage of dissolved organic carbon (DOC) through the landscape. The decomposition and subsequent leaching of organic litter in bogs, forests and wetlands are the principal sources of DOC in the terrestrial landscape. Production is mediated by several physical and biogeochemical factors, such as deposition of nitrates and sulphates from the atmosphere, moisture and temperature. The rate of export of terrestrial DOC is determined by the rate of production combined with the rate of sorption by mineral soils, and the availability of pathways for water through the landscape. Evans *et al.*¹ correlate an increase in hydrological DOC concentrations in the United Kingdom during 1988–2003 with the decreased deposition of sulphates in the form of acid rain in that period.

northwestern Canada and Alaska³. Although increased DOC concentrations in hydrological systems have been observed before^{4,5}, there has been no clear agreement on whether this constitutes a trend.

Where clear evidence of temporal changes in the concentration and/or export of DOC does exist, it is often difficult to provide convincing explanations. Many factors can influence its production and transport, and Evans and colleagues¹ review a range of hypotheses to explain their measured increases. First, they consider increased rates of decomposition in organic litter, and especially in peat soils, associated with changes in moisture and temperature⁶. This process could trigger an enzymatic 'latch' that would release more DOC⁷. Second, there is the possible stimulation of plant primary production by elevated atmospheric carbon dioxide⁸. Lastly, the authors examine the effect of changes in hydrological pathways.

After concluding that no proposed hypothesis is particularly convincing, Evans *et al.* posit a new one¹: that the increase in DOC concentration has been caused by the decrease in sulphur deposition that has occurred in northern Europe and North America over the past 20 years, as clean-air legislation has taken effect. The anthropogenic sulphur deposition is estimated to have declined by 41% between 1998 and 2003 in the stream and lake catchments that Evans and colleagues have studied⁹.

The mechanisms through which sulphur might influence DOC concentrations in this scheme are primarily chemical, rather than biological. Increased sulphate concentrations and raised pH in soil pore water have been

shown experimentally to decrease the mobilization of DOC in soils. To test this sulphur hypothesis, the authors present a correlation analysis of the relationship between the temporal trends in DOC concentrations and sulphate deposition.

But the origin and interactions of DOC in an entire landscape is a complex issue, and what is observed at the outflow of hydrological catchment areas is the combined result of many processes that occur upstream. In nine small boreal lakes in Canada, for example, variations in ice-free DOC concentrations were found to be most strongly correlated with solar radiation and winter precipitation⁵. Increases in sulphate deposition and lake acidity were not correlated with DOC concentration.

The problem is that in these types of analysis there is very little control on the experiments that nature is providing. Studies are plagued by the influence of extraneous variables, and the cause of changes in some of these variables is unknown. What is certain is that, at the heart of any DOC increase, there has to be an increase either in net DOC production in terrestrial ecosystems or in the leaching of DOC from them. Evans *et al.* demonstrate that in their catchment areas there has been no increase in runoff, thus focusing attention on the second option.

It is true that, during the period of their study, temperatures were increasing and sulphate deposition was decreasing, both in North America¹⁰ and in western Europe¹¹. But the same sources show that deposition of nitrates was increasing at the same time. This point is significant because nitrogen production can improve ecosystem productivity, resulting in

increased litter production, and can at the same time influence the rate of carbon mineralization in litter and soil. Elevated DOC concentrations in the Hudson River in the northeastern United States have been attributed to increased DOC production from forest soils subjected to higher nitrogen deposition, as well as to decreased degradability of DOC once in the river system¹². Similarly, it has been estimated that soils in England and Wales have lost 4 million tonnes of carbon each year from 1978 to 2003 because of climate change¹³. At least part of this loss might be as increased DOC transport to aquatic ecosystems.

The sulphur hypothesis could be tested. There are several regions in the developing world where sulphate deposition is rapidly increasing. If Evans and colleagues' hypothesis is correct, this should lead to a decrease in DOC stream concentration, once changes in the volume and pathways of water flow are taken into account.

Although the increase in DOC export in some regions is clear, studies showing that the trend is universal, and establishing its causes, are still wanting. Chemical analyses of the DOC in rivers and lakes or analysis of its ¹⁴C and ¹³C isotopic signatures might help in pinpointing its age and origin. An issue of more global import is whether there will be large changes in the concentration and export of DOC in rivers undergoing major changes as a result of the changing climate, such as the melting of permafrost. This might release large amounts of DOC², or alternatively allow greater water movement through mineral soils, and thus increase DOC retention through sorption³. Studies such as that of Evans and colleagues¹ are valuable contributions to the debate on rising DOC, but we are far from reaching definitive answers.

Nigel Roulet and Tim R. Moore are in the McGill School of Environment and the Department of Geography, McGill University, Montreal, Quebec H3A 2K6, Canada.
e-mail: nigel.roulet@mcgill.ca

- Evans, C. D., Chapman, P. J., Clark, J. M., Monteith, D. T. & Cresser, M. S. *Glob. Change Biol.* **12**, 2044–2053 (2006).
- Frey, K. E. & Smith, L. *Geophys. Res. Lett.* **32**, L09401 (2005).
- Striegl, R., Aiken, G., Dornblaser, M., Raymond, P. & Wickland, K. *Geophys. Res. Lett.* **32**, L21413 (2005).
- Evans, C., Monteith, D. & Copper, D. *Environ. Pollut.* **137**, 55–71 (2005).
- Hudson, J., Dillon, P. & Somers, K. *Hydrol. Earth Syst. Sci.* **7**, 390–398 (2003).
- Freeman, C., Evans, C. & Monteith, D. *Nature* **412**, 785 (2001).
- Freeman, C., Ostle, N. & Kang, H. *Nature* **409**, 149 (2001).
- Freeman, C. *et al.* *Nature* **430**, 195–198 (2004).
- Davies, J. J. L., Jenkins, A., Monteith, D. T., Evans, C. D. & Cooper, D. M. *Environ. Pollut.* **137**, 27–39 (2005).
- Meteorological Service of Canada *Canadian National Atmospheric Chemistry (NAtChem) Precipitation Chemistry Database* www.msc-smc.ec.gc.ca/natchem/precip/index_e.html (2006).
- Norwegian Pollution Control Authority *State of the Environment Norway* www.environment.no/templates/TopPage_3142.aspx (2006).
- Findlay, S. *Front. Ecol. Environ.* **3**, 133–137 (2005).
- Bellamy, P. H. *et al.* *Nature* **437**, 245–248 (2005).

BRIEF COMMUNICATIONS

Upper Palaeolithic infant burials

Decorations on the bodies of newborns indicate that they were probably important in their community.

Several adult graves from the Stone Age (Upper Palaeolithic period) have been found but child burials seem to be rare, which has prompted discussion about whether this apparently different treatment of infants could be significant^{1,2}. Here we describe two recently discovered infant burials from this period at Krems-Wachtberg in Lower Austria, in which the bodies were covered with red ochre and decorated with ornaments and were therefore probably ritually buried. These findings indicate that even newborns were considered to be full members of these hunter–gatherer communities about 27,000 years ago.

The Upper Palaeolithic sites in eastern Austria have been reinvestigated over the past decade. The loess sequences around Krems have been of particular interest, with surveys, test trenches and drilling-core analysis yielding insight into settlement patterns between the rivers Danube and Krems. Excavations at Krems-Hundssteig³ and Krems-Wachtberg⁴ (see supplementary information) have provided detailed information about the spatial organization of these camps.

In the first campaign at Krems-Wachtberg, an extraordinarily well-preserved living floor, radiocarbon dated at $26,580 \pm 160$ years before present (specimen Poz-1290) was found. It contained items of Gravettian (Upper Palaeolithic) culture, such as lithic artefacts and ornaments, as well as faunal remains, charcoal, ochre and a fired piece of clay bearing a human finger-print⁵. A double infant burial was discovered in 2005 ('burial 1') and a single infant burial in 2006 ('burial 2').

Each burial was recovered as a block, which was analysed by computer tomography and laser scanning of the surface layers. Burial 1 was in a 40-cm-long pit and the bodies were overlaid with an adult mammoth scapula, supported by part of a tusk. The two skeletons, which were very well preserved, were embedded in red ochre; individual A (Fig. 1a, bottom) was decorated with more than 30 ivory beads (shown enlarged in Fig. 1c). The developmental stage of a deciduous incisor of individual B (Fig. 1a, top) allowed the estimation of the age at death to be perinatal (ninth to tenth lunar month). The equal lengths of both right femora indicate that the newborns were the same age at death; their contemporaneous burial suggests that they were twins. The auditory ossicles of individual A (Fig. 1a, bottom) were also recovered.

Burial 2 (Fig. 1b), located about 1 metre north of burial 1, is of a single individual. From the degree of mineralization of the upper incisors, this infant must have died at 0–3 months after birth. The subsidence of the cultural layer above the pits provides evidence that all three infants were buried at the start of the settlement at the site.

The technology, the structure of the camp, and the evidence of ritual activities (use of red ochre, ornaments and the covering of the twin burial with a mammoth scapula) at Krems-Wachtberg attest to its close affinity with the South Moravian sites Dolní Věstonice and Předmostí⁶.

Nothing comparable to these burials of such young Upper Palaeolithic individuals has been

found before. They therefore expand the debate about Gravettian ritual(s) and add to the sparse sample of Palaeolithic human remains found so far in Europe⁷. Moreover, these well-preserved fossils of extremely young individuals will contribute important evidence to the study of the ontogeny of early modern humans.

Thomas Einwögerer*, **Herwig Friesinger***, **Marc Händel***, **Christine Neugebauer-Maresch***, **Ulrich Simon***, **Maria Teschler-Nicola†**

*Prehistoric Commission of the Austrian Academy of Sciences, 1010 Vienna, Austria
e-mail: christine.neugebauer-maresch@oeaw.ac.at

†Department of Anthropology, Natural History Museum Vienna, 1010 Vienna, Austria

1. Trinkaus, E. & Svoboda, J. *Early Modern Human Evolution in Central Europe. The People of Dolní Věstonice and Pavlov* (Oxford Univ. Press, 2006).
2. Zilhão, J. & Trinkaus, E. *Portrait of the Artist as a Child* (Trabalhos de Arqueologia 22) (Instituto Português de Arqueologia, Lisboa, 2002).
3. Neugebauer-Maresch, C. *Preistoria Alpina* **39**, 165–173 (2003).
4. Einwögerer, T. *Die Jungpaläolithische Station auf dem Wachtberg bei Krems, NÖ* (Mitt. Prähist. Komm. Österr. Akad. Wiss. 34) (Akademie der Wissenschaften, Wien, 2000).
5. Svoboda, J., Neugebauer-Maresch, C., Králík, M., Einwögerer, T. & Novotný, V. *Přehled Výzkumů* **45**, 256–259 (2004).
6. Svoboda, J. *Přehled Výzkumů* **46**, 63–82 (2004).
7. Teschler-Nicola, M. & Trinkaus, E. *J. Hum. Evol.* **40**, 451–465 (2001).

Supplementary information accompanies the communication on Nature's website.

Received 12 August; accepted 24 October 2006.

Competing financial interests: declared none.

doi: 10.038/444285a



Figure 1 | Upper Palaeolithic burial site at Krems-Wachtberg in eastern Austria. **a**, Double burial of two newborns (burial 1, excavated in 2005). These are in a flexed position on their left side, with their skulls oriented to the north and facing east. More than 30 ivory beads were found near the pelvis of individual A (at the bottom of the photo; enlarged in **c**). Both were embedded in a thick layer of red ochre and covered by the scapula of

a mammoth. **b**, Burial of an infant embedded in a thick layer of red ochre (burial 2, excavated in 2006) in a flexed position on its right side; the skull is oriented to the south. (Photographs: Natural History Museum of Vienna.) **c**, Detail of ivory beads associated with individual A in burial 1 at Krems-Wachtberg. (Photograph: Prehistoric Commission, Austrian Academy of Sciences.) Scale rules marked in 5 cm (**a,b**) and 1 cm (**c**).

MATERIALS

Carbon nanotubes in an ancient Damascus sabre

The steel of Damascus blades, which were first encountered by the Crusaders when fighting against Muslims, had features not found in European steels — a characteristic wavy banding pattern known as damask, extraordinary mechanical properties, and an exceptionally sharp cutting edge. Here we use high-resolution transmission electron microscopy to examine a sample of Damascus sabre steel from the seventeenth century and find that it contains carbon nanotubes as well as cementite nanowires. This microstructure may offer insight into the beautiful banding pattern of the ultrahigh-carbon steel created from an ancient recipe that was lost long ago.

It is believed that Damascus blades were forged directly from small cakes of steel (named 'wootz') produced in ancient India. A sophisticated thermomechanical treatment of forging and annealing was applied to these cakes to refine the steel to its exceptional quality. However, European bladesmiths were unable to replicate the process, and its secret was lost at about the end of the eighteenth century. It was unclear how medieval blacksmiths would have overcome the inherent brittleness of the plates of cementite (Fe_3C , a mineral known as cohenite) that form in steel with a carbon content of 1–2 wt%, as well as how the steel's characteristic banding could have arisen from these plates.

Mechanical processing at the appropriate temperature can cause the steel's microstructure to become fine-grained, and superplastic behaviour is induced at higher temperatures¹. Small additions of the elements vanadium, chromium, manganese, cobalt, nickel and others are known to facilitate the formation of cementite bands during thermal cycling at temperatures below the cementite formation temperature² (about 800 °C). Moreover, the Damascene steel contains rare-earth elements and shows evidence of cementite nanowires in its microstructure^{3–5}.

Using high-resolution transmission electron microscopy, we have now also detected carbon nanotubes in a specimen taken from a genuine Damascus sabre (sabre no.10 (ref. 6); sample kindly provided by E. J. Kläy of Berne Historical Museum, Switzerland) produced by the famous blacksmith Assad Ullah in the seventeenth century. Its microstructure has been investigated previously⁴, but the nanotubes only become apparent (Fig. 1) after dissolution of the sample in hydrochloric acid (for methods, see supplementary information). Some remnants (Fig. 1c) show evidence of incompletely dissolved cementite nanowires³, indicating that these wires could have been encapsulated

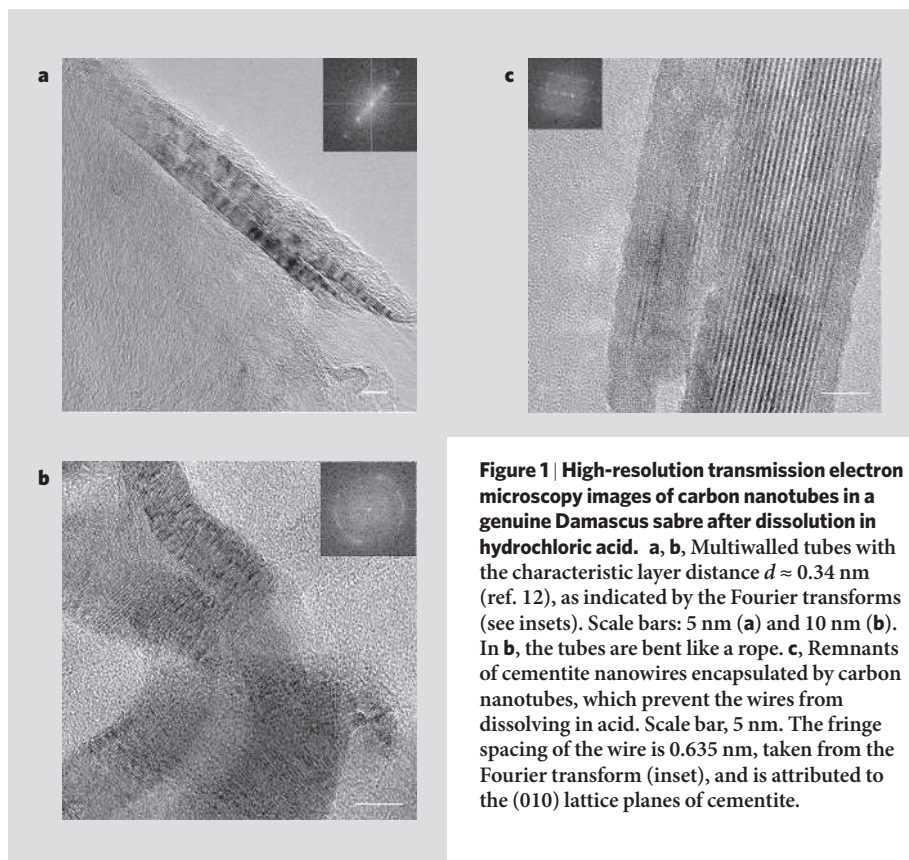


Figure 1 | High-resolution transmission electron microscopy images of carbon nanotubes in a genuine Damascus sabre after dissolution in hydrochloric acid. **a, b**, Multiwalled tubes with the characteristic layer distance $d \approx 0.34$ nm (ref. 12), as indicated by the Fourier transforms (see insets). Scale bars: 5 nm (**a**) and 10 nm (**b**). In **b**, the tubes are bent like a rope. **c**, Remnants of cementite nanowires encapsulated by carbon nanotubes, which prevent the wires from dissolving in acid. Scale bar, 5 nm. The fringe spacing of the wire is 0.635 nm, taken from the Fourier transform (inset), and is attributed to the (010) lattice planes of cementite.

and protected by the carbon nanotubes⁷.

Nanotubes can be formed catalytically⁸, as well as from hydrocarbons inside micropores at reduced temperatures⁹. We suggest therefore that our finding could link the distinctive banding seen in Damascus blades with 'impurities' contained in the steel^{2,4}. By empirically optimizing their blade-treatment procedure, craftsmen ended up making nanotubes more than 400 years ago.

According to an early report on Indian wootz production¹⁰, particular ingredients were mandatory — such as wood from *Cassia auriculata* and leaves of *Calotropis gigantea*, and ores taken from particular mines in India. The diminishing supply of some of these ores during the eighteenth century may have prevented bladesmiths, who would not have been aware of the importance of these alloying elements, from practising their ancient recipes.

Thermal cycling and cyclic forging cause catalytic elements to segregate gradually into planar arrays parallel to the forging plane¹¹. These elements may give rise to the growth of carbon nanotubes, which in turn initiate formation of cementite nanowires and coarse cementite particles. As the nanoscale structure of Damascus steel emerges, a refined

interpretation of its remarkable mechanical properties should become possible.

M. Reibold*†, **P. Paufler***, **A. A. Levin***, **W. Kochmann‡**, **N. Pätzke***, **D. C. Meyer***

*Institut für Strukturphysik, †Triebenberg Laboratory, Technische Universität Dresden, 01062 Dresden, Germany

e-mail: paufler@physik.tu-dresden.de

‡Krüllsstraße 4b, 06766 Wolfen, Germany

1. Wadsworth, J. *MRS Bull.* **27**, 980–987 (2002).
2. Verhoeven, J. D. *Steel Res.* **73**, 356–365 (2002).
3. Kochmann, W. et al. *J. Alloys Comp.* **372**, L15–L19 (2004).
4. Levin, A. A. et al. *Crystal Res. Technol.* **40**, 905–916 (2005).
5. Reibold, M., Levin, A. A., Meyer, D. C., Paufler, P. & Kochmann, W. *Int. J. Mater. Res.* **97**, 1172–1182 (2006).
6. Zschokke, B. *Rev. Metallurg.* **21**, 635–669 (1924).
7. Golberg, D. et al. *Acta Mater.* **54**, 2567–2576 (2006).
8. Ni, L. et al. *Carbon* **44**, 2265–2272 (2006).
9. Chernozatonskii, L. A. et al. *Carbon* **35**, 749–753 (1997).
10. Schwarz, C. *Stahl u. Eisen* **21**, 209–211 (1901).
11. Verhoeven, J. D., Pendray, A. H. & Dauksch, W. *JOM* **56**, 17–20 (2004).
12. He, C., Zhao, N., Shi, C., Du, X. & Li, J. *Mater. Chem. Phys.* **97**, 109–115 (2006).

Supplementary information accompanies this communication on Nature's website.

Received 24 July; accepted 25 October 2006.

Competing financial interests: declared none.
doi:10.1038/444286a

BRIEF COMMUNICATIONS ARISING online
 ▶ www.nature.com/bca see Nature contents.

EARTH SCIENCE

Palaeo-altimetry of Tibet

Arising from: D. B. Rowley & B. S. Currie *Nature* **439**, 677–681 (2006).

The determination of palaeo-elevation has emerged in the past 15 years as an important tool for constraining physical processes that govern the formation of mountain belts. Rowley and Currie¹ report palaeo-elevations for the Lunpola basin within the Tibetan plateau and claim that these elevations are incompatible with 'mantle-thickening models' for mountain formation. We show here that their data do not support this conclusion and, indeed, are consistent with its opposite. The Tibetan plateau could have risen by a kilometre or more as its dense lower lithosphere sank into the underlying mantle.

If continental lithosphere is thickened to form a mountain belt, the lower part of the lithosphere, which is inherently unstable because it has higher density than its underlying mantle, may abruptly sink and be replaced by mantle that is less dense^{2,3}. Depending on the thickness of lithosphere that is replaced, the Earth's surface can rise by 1,000 m to perhaps 2,500 m (refs 2,3). (In this communication, all references to elevation are to the height of the land surface above sea level; change in sea level between 35 Myr and the present is negligible in this context.) The essence of Rowley and Currie's argument is that, because their inferred elevations for the Lunpola basin at about 35 Myr (4,850 m (+1,630/–1,435 at 95% confidence), 4,260 m (+1,480/–1,420), and 4,050 m (+1,420/–1,220)) are similar to its present-day elevation (4,567–4,718 m), the surface of the Tibetan plateau cannot have risen appreciably after 35 Myr, and that convective removal of lower lithosphere could not have occurred.

Rowley and Currie's¹ argument assumes that the elevation history of the Lunpola basin, which forms only a very small part of the Tibetan plateau, is representative of the whole. The position of the basin renders that

assumption doubtful: it lies close to the northern edge of the Lhasa block, which was part of an Andean margin through late Cretaceous and early Tertiary time, and probably achieved a high elevation much earlier than the rest of the plateau^{4–6}.

Rowley and Currie's¹ argument also neglects the influence of crustal thinning on the surface height of the plateau since 35 Myr ago. Extensional faulting began on the plateau at about 8–15 Myr^{3,7,8}, and summation of moment tensors of earthquakes suggests that roughly half of the present-day rate of east–west extension in Tibet represents vertical thinning^{9,10}. Because of isostasy, crustal thinning causes the surface elevation to drop. The present rate of thinning is about 4×10^{-9} per yr; if this rate had operated for the past 8–15 Myr, then the crust would have thinned by 2.5–5 km and, in the absence of other processes, the height of the land surface would have decreased by about 500–1,000 m (refs 2, 3). Because crustal thinning extracts gravitational potential energy from the plateau^{2,3}, the present-day rate of thinning is probably lower than the average rate, so the total decrease in surface height may have been greater than we estimate.

The average of Rowley and Currie's¹ three estimates of palaeo-elevation, at 35 Myr, for sites in the Lunpola basin is about 250 m lower than the present-day elevation of the basin and, at 8–15 Myr, surface heights in the plateau were higher than they are now by about 500–1,000 m. Thus, ignoring uncertainties, Rowley and Currie's¹ palaeo-altitudes are consistent with a rise of the Lunpola basin by about 750–1,250 m at some time between 35 ± 5 Myr and the onset of crustal thinning at around 8–15 Myr. Accounting for departures from assumptions in Rowley and Currie's estimates¹ (such as differences in initial $\delta^{18}\text{O}$, in paths of

moisture transport, in evaporation of precipitation, and so on) could make the uncertainties larger than quoted, and permit greater surface uplift. Thus, Rowley and Currie's results do not disprove 'mantle-thickening models' for the formation of the plateau: indeed, when combined with the effect of crustal thinning on the surface elevation of the plateau in the past 15 Myr, the data offer support for the idea that convective removal of mantle lithosphere contributed to the high surface elevation of the present-day Tibetan plateau.

Peter Molnar*, Gregory A. Houseman†, Philip C. England‡

*Department of Geological Sciences and Cooperative Institute for Research in Environmental Science (CIRES), University of Colorado, Boulder, Colorado 80309-0399, USA
e-mail: molnar@cires.colorado.edu

†Institute of Geophysics and Tectonics, School of Earth and Environment, University of Leeds, Leeds LS2 9JT, UK

‡Department of Earth Sciences, Oxford University, Oxford OX1 3PR, UK

1. Rowley, D. B. & Currie, B. S. *Nature* **439**, 677–681 (2006).
2. England, P. & Houseman, G. J. *Geophys. Res.* **94**, 17561–17579 (1989).
3. Molnar, P., England, P. & Martinod, J. *Rev. Geophys.* **31**, 357–396 (1993).
4. England, P. C. & Searle, M. *Tectonics* **5**, 1–14 (1986).
5. Murphy, M. A. et al. *Geology* **25**, 719–722 (1997).
6. Kapp, P., Yin, A., Harrison, T. M. & Ding, L. *Geol. Soc. Am. Bull.* **117**, 865–878 (2005).
7. Harrison, T. M., Copeland, P., Kidd, W. S. F. & Yin, A. *Science* **255**, 1663–1670 (1992).
8. Blisniuk, P. M. et al. *Nature* **412**, 628–632 (2001).
9. Molnar, P. & Chen, W.-P. *J. Geophys. Res.* **88**, 1180–1196 (1983).
10. Molnar, P. & Lyon-Caen, H. *Geophys. J. Int.* **99**, 123–153 (1989).

doi:10.1038/nature05368

extreme: for example, present-day Java is comparably 'Andean' but elevations in the back-arc extend to below sea level, despite contractional deformation in this region. Although evidence from Cretaceous–early Tertiary shortening in central Tibet has provided insight into the regional extent of pre-collision shortening⁵, the magnitude of crustal thickening, and its corresponding effect on Tibetan palaeo-elevations, is unresolved.

Continuous convergence of India and Asia since collision began requires either storage of intrinsically negatively buoyant mantle lithosphere above the asthenosphere, or continuous advective removal of mantle lithosphere from both the Indian and Asian sides of the orogen into the underlying mantle, or some

EARTH SCIENCE

Rowley & Currie reply

Replying to: P. Molnar, G. A. Houseman & P. C. England *Nature* **444**, doi:10.1038/nature05368 (2006).

Molnar *et al.*¹ question our conclusion on the role of convective destabilization of thickened mantle lithosphere in determining the surface elevation history of the Tibetan plateau. The primary argument depends on our interpretation^{2,3} of oxygen-isotope-based estimates of palaeo-elevation of the Lunpola basin, a locality in the centre of the plateau.

Molnar *et al.*¹ raise the questions of whether

the Lunpola basin is representative of the plateau and whether our data are incompatible with an increase of 1 km to 2.5 km in the surface elevation⁴. Regarding their first point, it is not known how representative the Lunpola basin is of the central Tibetan plateau, nor whether it was topographically high before the date of our oldest samples. However, Molnar *et al.* may be carrying the Andean analogy to

intermediate response. Results on convective destabilization of the mantle lithosphere cited by Molnar *et al.*¹ prescribe mantle lithosphere thickening and convective destabilization, but other responses are possible⁶, and our data² seem to support other such responses.

Regarding the second point of Molnar *et al.*¹, we believe that they overinterpret our results and neglect aspects that bias the data in ways that cause the isotopes to yield lower palaeo-elevation estimates, particularly for the lacustrine limestone of the Middle Niubao and Dingqing Formations. The oxygen-isotope signature of soil and lacustrine carbonates is primarily controlled by elevation and evaporation. The relationship predicted between isotope composition and elevation is, in turn, largely controlled by the temperature and, to a much lesser degree, relative humidity of the moisture-laden air masses before orographic lifting³.

We gave two different estimates of uncertainty², which may have led to the confusion.

One estimate reflects uncertainty in the isotopic compositions and uncertainties in the equilibrium temperature of carbonate formation, and in the isotopic composition of the low-altitude source from which we derive $\Delta(\delta^{18}\text{O})$. The second estimate of uncertainty is in the relation between $\Delta(\delta^{18}\text{O})$ and elevation³. This estimate reflects variations in potential starting temperature and relative humidity for areas of low latitude. Our palaeo-altitude estimates² are based on modern, low-latitude temperatures and relative humidities³.

Eocene, Oligocene and Miocene palaeo-climates were probably warmer than at present³, which would increase our estimated palaeo-altitudes by about 1,000 m and be compatible with a progressive decrease in elevation over the past 35 ± 5 Myr; our estimates are therefore minima. Similarly, evaporative re-enrichment of surface waters decreases estimates of $\Delta(\delta^{18}\text{O})$ and hence of surface elevation, probably affecting the Middle Niubao and Dingqing Formations results the most. Our $\Delta(\delta^{18}\text{O})$

values are more likely to be minimal estimates, as are the resulting palaeo-altitudes of the Lunpola region at the time of their deposition. We do not believe our data can resolve the modest decrease in surface elevation, if it occurs, associated with east–west extension.

D. B. Rowley*, B. S. Currie†

Department of Geophysical Sciences, University of Chicago, Chicago, Illinois 60637, USA

e-mail: rowley@geosci.uchicago.edu

† Department of Geology, Miami University, Oxford, Ohio 45056, USA

1. Molnar, P., Houseman, G. A. & England, P. C. *Nature* **444**, doi:10.1038/nature05368 (2006).
2. Rowley, D. B. & Currie, B. S. *Nature* **439**, 677–681 (2006).
3. Rowley, D. B., Pierrehumbert, R. T. & Currie, B. S. *Earth Planet. Sci. Lett.* **188**, 253–268 (2001).
4. Molnar, P., England, P. & Martinod, J. *Rev. Geophys.* **31**, 357–396 (1993).
5. Kapp, P., Yin, A., Harrison, T. M. & Ding, L. *Geol. Soc. Am. Bull.* **117**, 865–878 (2005).
6. Pysklywec, R. N., Beaumont, C. & Fullsack, P. *Geology* **28**, 655–658 (2000).

doi:10.1038/nature05369

**Cover illustration**

For many animals, such as the rat pictured here, smell and taste are crucial for sensing their environment.

Editor, *Nature*

Philip Campbell

Insights Publisher

Sarah Greaves

Insights Editor

Lesley Anson

Production Editor

Sarah Archibald

Senior Art Editor

Martin Harrison

Art Editor

Nik Spencer

Sponsorship

Claire Hines

Claudia Banks

Production

Susan Gray

Marketing

Katy Dunningham

Editorial Assistant

Laura Shaw

CHEMICAL SENSING

Taste and smell contribute critically to our experience of our environment, from the pleasure of eating to the formation of childhood memories. Despite this, research into chemical sensing historically took a back seat to vision and audition, challenged by the complexities of indefinable stimuli that are often not even consciously processed, and by difficulties in uncovering even the basic mechanisms of transduction and representation.

Then, in 1991, Buck and Axel identified the genes for odorant receptors, launching chemical-sensing research to the forefront. The reviews in this Insight highlight the progress made during the past fifteen years, and emphasize how the stage is set to tackle fundamental questions about the brain and behaviour.

New experimental tools, such as sophisticated methods for labelling, recording and manipulating neuronal activity and projections, have provided researchers with insight into the principles by which the taste and olfactory systems decode and represent environmental stimuli. An important goal that now seems attainable is to understand how this processing translates into, in some instances, innate behaviours and, in others, learned behaviours.

The findings reviewed here are not only remarkable intellectual achievements, but are also of great potential value to society. Understanding the cues and mechanisms through which humans and other animals are attracted to or repulsed by chemosensory stimuli holds promise for tackling several major problems facing modern human society — from the control of disease-bearing insects to causes of overeating.

We are pleased to acknowledge the financial support of our principal sponsor Givaudan and our supporting sponsors ChemCom and the National Institute on Deafness and Other Communication Disorders in producing this Insight. As always, *Nature* carries sole responsibility for all editorial content and peer review.

I-han Chou, Tanguy Chouard, John E. Spiro & Lesley Anson, Senior Editors

REVIEWS

288 The receptors and cells for mammalian taste

J. Chandrashekar, M. A. Hoon, N. J. P. Ryba & C. S. Zuker

295 Comparative chemosensation from receptors to ecology

C. I. Bargmann

302 Insects as chemosensors of humans and crops

W. van der Goes van Naters & J. R. Carlson

308 Pheromonal communication in vertebrates

P. A. Brennan & F. Zufall

316 Smell images and the flavour system in the human brain

G. M. Shepherd

nature
insight

The receptors and cells for mammalian taste

Jayaram Chandrashekar¹, Mark A. Hoon², Nicholas J. P. Ryba² & Charles S. Zuker¹

The emerging picture of taste coding at the periphery is one of elegant simplicity. Contrary to what was generally believed, it is now clear that distinct cell types expressing unique receptors are tuned to detect each of the five basic tastes: sweet, sour, bitter, salty and umami. Importantly, receptor cells for each taste quality function as dedicated sensors wired to elicit stereotypic responses.

Our sensory systems are responsible for generating an internal representation of the outside world, including its chemical (taste and olfaction) and physical (mechanical, sound, vision and temperature) features. In this review we examine recent advances in our understanding of the biology of taste, focusing on receptors, cells and the logic of taste coding at the periphery.

Taste is in charge of evaluating the nutritious content of food and preventing the ingestion of toxic substances. Sweet taste permits the identification of energy-rich nutrients, umami allows the recognition of amino acids, salt taste ensures the proper dietary electrolyte balance, and sour and bitter warn against the intake of potentially noxious and/or poisonous chemicals. In humans, taste has the additional value of contributing to the overall pleasure and enjoyment of a meal. Surprisingly, although we can taste a vast array of chemical entities, it is now generally accepted that, qualitatively, they evoke few distinct taste sensations: sweet, bitter, sour, salty and savoury (or umami). Although this repertoire may seem modest, it has satisfactorily accommodated the evolutionary need for an effective and reliable platform to help recognize and distinguish key dietary components.

The anatomical substrates and units of taste detection are taste-receptor cells (TRCs; Fig. 1). TRCs are assembled into taste buds, which are distributed across different papillae of the tongue and palate epithelium. How are the different tastes detected, and how is taste quality represented? In the simplest scenario, sweet, bitter, sour, salty and umami tastants would each be recognized by different cells expressing specialized receptors. Coding at the periphery could then rely on straightforward labelled lines (that is, independent sweet, bitter, sour, salty and umami signals) to transform tastant quality into neural signals (Fig. 2a). In an alternative view, and the prevailing model for the past two decades^{1–3}, TRCs were proposed to be broadly tuned across taste modalities (Fig. 2b, c). In this case, it would be expected that individual TRCs would express different families of taste receptors, and that tastant recognition would result from decoding of the combined activity of various classes of such broadly tuned TRCs (the ‘across-fibre pattern’ of coding)^{4,5}. The recent identification of cells and receptors mediating sweet, bitter, umami and sour taste (Figs 3, 4; Table 1) has generated powerful molecular tools that can be used to devise rigorous tests to distinguish between these models and establish the basis of taste coding at the periphery.

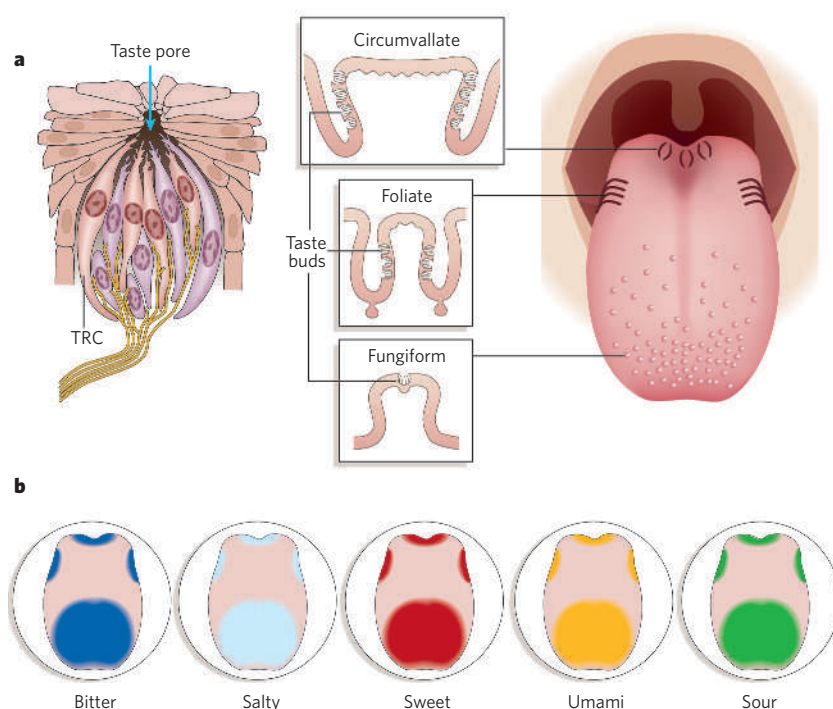


Figure 1 | Taste-receptor cells, buds and papillae. **a**, Taste buds (left) are composed of 50–150 TRCs (depending on the species), distributed across different papillae. Circumvallate papillae are found at the very back of the tongue and contain hundreds (mice) to thousands (human) of taste buds. Foliate papillae are present at the posterior lateral edge of the tongue and contain a dozen to hundreds of taste buds. Fungiform papillae contain one or a few taste buds and are found in the anterior two-thirds of the tongue. TRCs project microvilli to the apical surface of the taste bud, where they form the ‘taste pore’; this is the site of interaction with tastants. **b**, Recent molecular and functional data have revealed that, contrary to popular belief, there is no tongue ‘map’: responsiveness to the five basic modalities — bitter, sour, sweet, salty and umami — is present in all areas of the tongue^{6,8,9,32,78}.

¹Howard Hughes Medical Institute and Departments of Neurobiology and Neurosciences, University of California at San Diego, La Jolla, California 92093-0649, USA. ²National Institute of Dental and Craniofacial Research, National Institutes of Health, Bethesda, Maryland 20892, USA.

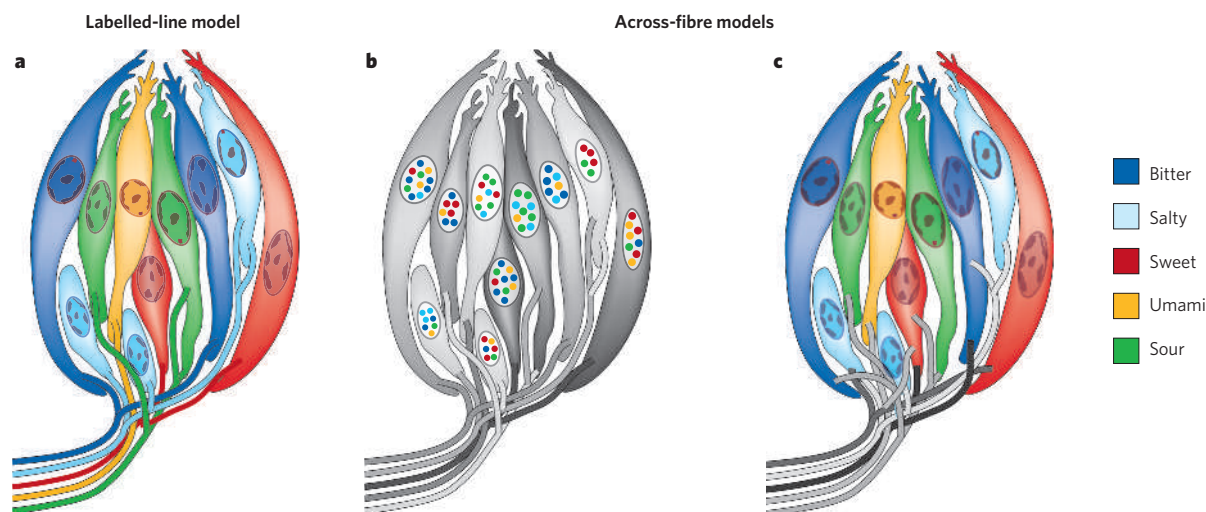


Figure 2 | Encoding of taste qualities at the periphery. There are two opposing views of how taste qualities are encoded in the periphery. **a**, In the labelled-line model, receptor cells are tuned to respond to single taste modalities — sweet, bitter, sour, salty or umami — and are innervated by individually tuned nerve fibres. In this case, each taste quality is specified by the activity of non-overlapping cells and fibres. **b, c**, Two contrasting models of what is known as the ‘across-fibre pattern’. This states that either individual TRCs are tuned to multiple taste qualities (indicated by various tones of grey and multicoloured stippled nuclei), and consequently the same afferent fibre carries information for more than one taste modality (**b**), or that TRCs are still tuned to single taste qualities but the same afferent fibre carries information for more than one taste modality (**c**). In these two models, the specification of any one taste quality is embedded in a complex pattern of activity across various lines. Recent molecular and functional studies in mice have demonstrated that different TRCs define the different taste modalities, and that activation of a single type of TRC is sufficient to encode taste quality, strongly supporting the labelled-line model.

Sweet taste

The sweetness of sugar and the pleasure it evokes are so familiar to us that they almost seem to be physical properties of sucrose rather than a representation of neuronal firing in the brain. This tight relationship between sensory quality, positive hedonic value and behavioural acceptance richly illustrates how sweet taste detection and perception evolved to help with the recognition of the most basic and fundamental sources of metabolic energy.

The attractive taste modalities, sweet and umami, are mediated by a small family of three G-protein-coupled receptors (GPCRs) — T1R1, T1R2 and T1R3 — that is distantly related to metabotropic glutamate, pheromone, extracellular-calcium sensing and γ -aminobutyric-acid type B receptors^{6–15}. These GPCRs assemble into either homodimeric or heterodimeric receptor complexes¹⁶, and are characterized by the presence of long amino-terminal extracellular domains that are believed to mediate ligand recognition and binding¹⁷.

The critical role of T1Rs in sweet taste detection and perception emerged from an ensemble of studies, including the characterization of T1R expression profiles, the analysis of naturally occurring sweet receptor mutants (and the identification of species-specific differences in sweet taste preferences), functional experiments in cell-based assays, and the generation of genetically modified mouse lines.

T1Rs are expressed in subsets of TRCs, and their expression pattern defines three cell types: TRCs co-expressing T1R1 and T1R3 (T1R1+3 cells), TRCs co-expressing T1R2 and T1R3 (T1R2+3 cells) and TRCs containing T1R3 alone⁸. What do these cells do? More than 30 years ago, genetic studies of sweet taste in mice identified a single principal locus that influences responses to several sweet substances^{18,19}. This locus, known as *Sac*, determines threshold differences in the ability of some strains to distinguish sucrose- and saccharin-containing solutions from water¹⁹. The *Sac* locus was recently shown by linkage analysis^{8,11–14,20} and genetic rescue⁸ to encode T1R3, thus implicating a member of the *T1r* gene family in sweet taste detection. Indeed, functional expression studies in heterologous cells revealed that T1R3 combines with T1R2 (T1R2+3) to form a sweet taste receptor that responds to all classes of sweet tastants, including natural sugars, artificial sweeteners, D-amino acids and intensely sweet proteins^{8,10}. These results validated the T1R2+3 heteromer as a sweet receptor, and suggested that T1R2+3 cells are the sweet-sensing TRCs (see below).

Humans and mice show some prominent differences in their ability to taste certain artificial sweeteners and intensely sweet proteins — for example, mice cannot taste aspartame or monellin. Notably, introduction of the human T1R2 receptor into mice significantly changes their sweet taste preferences to a human-like response profile¹⁵, proving that species differences in sweet taste sensitivity and selectivity are a direct reflection of T1R-sequence variation between species. How does a single receptor complex respond to such a wide range of sweet-tasting compounds, from simple six-carbon sugars to guanidinoacetic acids and even large peptides²¹ and polypeptides? Recently, biochemical studies of human, rodent and chimaeric human–rodent T1R2+3 receptors have shown that diverse classes of sweet-receptor ligands actually require different domains of the receptor complex for recognition^{22–24}, thus providing a simple solution to this puzzle. Together, these genetic, functional and biochemical studies have amply validated the role of the T1R2 and T1R3 subunits in sweet-tastant recognition, and demonstrated the importance of heteromerization in receptor function.

Definitive proof that T1R2+3 is the principal mammalian sweet taste receptor was obtained from studies of *T1r2*- and *T1r3*-knockout mice^{15,25} (Fig. 3). Homozygous mutants for either receptor subunit show a devastating loss of sweet taste — all behavioural and electrophysiological responses to artificial sweeteners, D-amino acids and low to moderately high concentrations (up to 300 mM) of natural sugars are abolished^{15,25}. However, these animals retain very small, albeit measurable, responses to very high concentrations of sugars. Importantly, a *T1r2*/*T1r3* double knockout completely eliminated these remaining sweet responses¹⁵, unequivocally demonstrating the essential role of T1Rs in all sweet taste detection and perception. Unexpected corroboration of the fundamental requirement of T1Rs for sweet taste came from the recent discovery that cats (all felidae from the common house kitten to the tiger) carry a naturally occurring deletion in their *T1r2* gene²⁶, providing a molecular explanation to the striking, and long-standing, observation that cats do not respond to sweets.

Umami taste

Most mammals are robustly attracted to the taste of a broad range of L-amino acids^{15,27–29}. In humans, however, just two amino acids — monosodium glutamate (MSG) and aspartate — evoke the unique savory sensation known as umami (whose Japanese characters can be

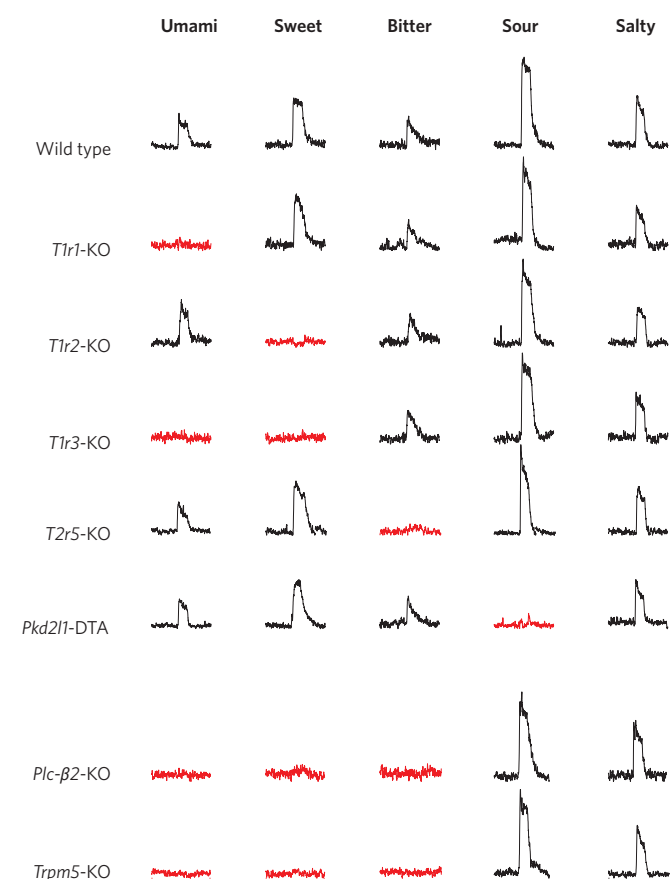


Figure 3 | Sweet, umami, bitter and sour are mediated by specific receptors and cells. The traces show recordings of tastant-induced activity in nerves innervating the tongue in wild-type and various gene-knockout (KO) mice or cell ablation studies (*Pkd2l1*-DTA). T1R1+3 functions as the umami receptor, T1R2+3 is the sweet receptor, T2Rs are bitter receptors (T2R5 is a high-affinity cycloheximide receptor), PKD2L1 is a candidate sour receptor, and PLC- β 2 is the effector and TRPM5 the transduction channel of sweet, umami and bitter pathways. Note the extraordinarily specific taste deficits (red traces) in each genetically altered mouse line. *Pkd2l1*-DTA refers to animals expressing diphtheria toxin in PKD2L1 cells.

translated as 'delicious flavour'³⁰, perhaps best exemplified in western cuisine by the taste of meaty broths. A salient feature of amino-acid taste in animals, and umami taste in humans is their impressive potentiation by purine nucleotides (such as IMP and GMP)³¹. This feature has been cleverly commandeered by the food industry as a means of enhancing the flavour of a wide range of products, and was expected to be a biochemical hallmark of the authentic umami receptor.

Cell-based expression studies have shown that the rodent T1R1 and T1R3 GPCRs combine to form a broadly tuned L-amino-acid receptor⁹. These results validated T1R1+3 as an amino-acid taste receptor, and the T1R1+3-expressing cells as candidate umami-sensing cells. Interestingly, in cell-based assays, the human T1R1+3 complex functions as a much more specific receptor, responding selectively to monosodium glutamate and aspartate (as well as to the glutamate analogue L-AP4), with sensitivity that recapitulates human psychophysical thresholds for umami taste^{9,10}. In addition, as would be predicted for the genuine umami receptor, both the rodent and human T1R1+3 heterodimers showed strong potentiation in response to purine nucleotides^{9,10}.

Final proof that T1R1+3 functions *in vivo* as the amino-acid (umami) taste receptor was obtained from the study of *T1r1*- and *T1r3*-knockout mice^{15,25} (Fig. 3). Homozygous mutants lacking either the T1R1 or T1R3 subunit showed an overwhelming loss of umami taste, including all responses to IMP and behavioural attraction to monosodium

glutamate and L-amino acids¹⁵ (but see also ref. 25). Together, these results firmly established the T1R1+3 heteromeric GPCR complex as the mammalian umami taste receptor and provided a striking example of heteromeric GPCRs radically altering their selectivity according to a combinatorial arrangement of subunits (sweet T1R2+3 versus umami T1R1+3). They also revealed that sweet and amino-acid (umami) taste — two chemosensory inputs that trigger behavioural attraction — share a common receptor repertoire and evolutionary origin.

Bitter taste

In contrast to sweet and umami taste, which evolved to recognize a limited subset of nutrients, bitter taste has the onerous task of preventing the ingestion of a large number of structurally distinct toxic compounds. Remarkably, despite the vastness of this repertoire, these compounds all evoke such a similar sensation that we simply know them as 'bitter'. These observations suggest that bitter taste receptors are probably encoded by a large family of genes, and that the bitter sensation evolved to allow recognition of a wide range of chemicals, but not necessarily to distinguish between them.

Bitter taste is mediated by a family of ~30 highly divergent GPCRs (the T2Rs)^{32,33}. T2R genes are selectively expressed in subsets of TRCs distinct from those containing sweet and umami receptors³², and are clustered in regions of the genome genetically linked to bitter taste in humans and mice^{32–35}. A large number of T2Rs have been shown to function as bitter taste receptors in heterologous expression assays^{36–40}, and several have distinctive polymorphisms that are associated with significant variations in sensitivity to selective bitter tastants in mice³⁶, chimpanzees⁴¹ and humans⁴².

Proof that T2Rs are necessary and sufficient for bitter taste came from knockout and misexpression studies in mice. On the one hand, animals lacking a specific T2R (for example, T2R5, the candidate cycloheximide receptor), exhibited a marked and selective loss of their ability to taste the cognate bitter compound⁴³ (Fig. 3). On the other hand, mice engineered to express the human candidate receptors for PTC (phenylthiocarbamide) and salicin, two bitter substances that mice do not normally respond to, became vigorously averse to these two chemicals⁴³. These results demonstrated that T2Rs are necessary and sufficient for selective responses to bitter tastants, and validated T2Rs and T2R-expressing cells as the *in vivo* mediators of bitter taste detection and perception. In addition, the fact that the bitter taste responses of mice can be humanized by introducing human taste receptors illustrated an important feature of T2Rs and bitter taste: selectivity and sensitivity differences to bitter compounds between species is likely to be a reflection of sequence differences in their respective T2R repertoires^{44,45}.

A remarkable feature of bitter-receptor biology was exposed by the discovery that most, if not all, T2Rs are expressed in the same TRCs³². This implied that individual T2R-expressing cells may function as broadly tuned sensors for all bitter chemicals but might have very limited, if any, discrimination. In fact, it would not be unreasonable to imagine that although animals must be able to detect many bitter compounds, they have no need to distinguish between them qualitatively. Indeed, recent studies in mice have confirmed that T2R-expressing cells operate as universal bitter sensors^{43,46}, and that, although mice and rats can distinguish differences in intensity between bitter tastants, they are incapable of discriminating between them⁴⁷. Of course, it would be unreasonable to expect that different bitter TRCs express the same T2R proteins at identical levels, and thus individual bitter-sensing cells can be predicted to vary in their sensitivity to bitter tastants but still be able to respond to the full repertoire.

Signalling downstream of T1Rs and T2Rs

Signalling cascades downstream of taste receptors have been the subject of intense speculation over the years, with most models hypothesizing a surprising diversity of pathways and strategies^{48–50}. This proposed complexity contrasted sharply with the demonstrated simplicity of the signalling pathways of other senses, such as olfaction, in which hundreds of distinct receptors share an identical transduction cascade⁵¹.

Sure enough, recent results have demonstrated that the receptors for sweet, bitter and umami taste, although expressed in separate subsets of cells⁸, all signal through a common pathway to transduce tastant recognition into cell activation⁴⁶ (but see also ref. 52).

The current data suggest that tastant binding to T1Rs or T2Rs activates the heterotrimeric G proteins gustducin⁵³ or Gα₁₂ (ref. 54) leading to the release of the Gβγ subunits^{46,55} and the subsequent stimulation of phospholipase Cβ2 (PLC-β2)^{46,56}. Activation of PLC-β2 hydrolyses phosphatidylinositol-4,5-bisphosphate to produce the two intracellular messengers inositol-1,4,5-trisphosphate and diacylglycerol, and ultimately leads to the gating of the taste-transduction channel (the transient receptor potential (TRP) protein TRPM5)^{46,57}. As expected from this model, mouse knockouts of gustducin^{58,59}, PLC-β2 (refs 46, 60) or TRPM5 (refs 46, 52) have major deficits in sweet, umami and bitter tastes (Fig. 3). Importantly, salty and sour tastes remain unimpaired in all cases⁴⁶, demonstrating that these two modalities use a different signalling pathway and operate independently of sweet, umami and bitter tastes.

Are other pathways important in the detection and perception of sweet, bitter and umami tastes? We would expect signalling molecules representing many different transduction cascades to be present in most types of cell, including TRCs^{56,61–66}. However, their mere presence does not imply that they must be involved in taste transduction⁵⁰. Comprehensive physiological and genetic studies will be required to assess what role, if any, they have in taste signalling. It would be interesting if second messengers from a number of pathways modulate taste signals, both at the receptor and downstream levels, and thus provide a platform to shape taste responses as a function of various cues and cellular states. Intriguingly, activation of the TRPM5 transduction channel was recently shown to be strongly temperature dependent⁶⁷, at a range within the normal function of TRCs (15–35 °C). Talavera and colleagues⁶⁷ proposed that this property of the channel may underlie some of the effects of temperature on taste detection and, ultimately, perception⁶⁷. It would be rewarding to engineer animals expressing TRPM5 channels with modified temperature profiles and determine the behavioural and physiological impact of such changes on taste responses.

Salt and sour tastes

A number of studies have suggested that salty and sour tastants modulate taste-cell function by direct entry of Na⁺ and H⁺ through specialized membrane channels on the apical surface of the cell. In the case of salt, TRC activation is believed to be mediated at least in part by the entry of Na⁺ through amiloride-sensitive Na⁺ channels^{68,69}. However, the identity of the salt 'receptor' remains speculative and highly controversial^{68,70}.

A broad range of cell types, receptors and mechanisms have been proposed to be responsible for sour taste. These include the activation of

hyperpolarization-activated cyclic-nucleotide-gated (HCN) channels⁷¹, acid-sensing ion channels (ASICs)⁷², potassium (K2P) channels^{73,74} and H⁺-gated calcium channels⁷⁵, as well as the involvement of Na⁺/H⁺ exchangers⁷⁶ and acid inactivation of K⁺ channels⁷⁷. However, recent genetic and functional studies have greatly simplified the quest for the sour receptor by demonstrating that a member of the TRP ion-channel family, PKD2L1, demarcates sour-sensing TRCs⁷⁸. PKD2L1 is selectively expressed in a population of TRCs distinct from those mediating sweet, umami and bitter tastes^{78,79}, further substantiating the cellular segregation of taste modalities at the periphery.

Proof that PKD2L1-expressing cells function as the acid receptors in the taste system came from conclusive genetic-ablation experiments. The targeting of diphtheria toxin to PKD2L1-expressing cells of the tongue produced animals with a specific and total loss of sour taste⁷⁸ (Fig. 3). These results validated PKD2L1 TRCs as the sole acid-sensing cells and implicated the PKD2L1 ion channel as the candidate component of the sour taste (pH) receptor^{78,80}. The further demonstration that these sour-deficient mice have normal salt responses indicates that salt taste also must be mediated by an independent population of TRCs (see Fig. 3 and below).

Although sweet, umami and bitter sensing are primarily required in the taste system, acid sensing is also important in a number of other processes, including the monitoring of CO₂ levels in the blood⁸¹ and the internal state of the cerebrospinal fluid and brain⁸². Consequently, it may be predicted that PKD2L1 might also function in other physiological settings. Indeed, Huang and colleagues⁷⁸ showed that PKD2L1 is expressed in a selective population of neurons contacting the central canal of the spinal cord that fire in response to minor changes in proton concentration. These results suggest that these neurons function as sentinels of cerebrospinal and ventricular pH, and bring forth a surprising unity in the cellular basis of pH sensing in very different physiological systems.

Taste coding at the periphery

Several electrophysiological and calcium-imaging-based studies in rats and mice have reported that individual TRCs are tuned to various taste modalities^{4,83–85} and have proposed that encoding of taste quality at the periphery must use an across-fibre pattern of activity (Fig. 2). However, the discovery that sweet, umami, bitter and sour (and, by extrapolation, salt) taste cells are segregated into non-overlapping populations expressing distinct receptors (Fig. 4) demands a revision of this model. We review three lines of investigation demonstrating that the TRCs defined by T1Rs, T2Rs and PKD2L1 function as highly dedicated sensors for sweet, umami, bitter and sour tastes, strongly arguing, instead, for a labelled-line model of coding across all taste modalities.

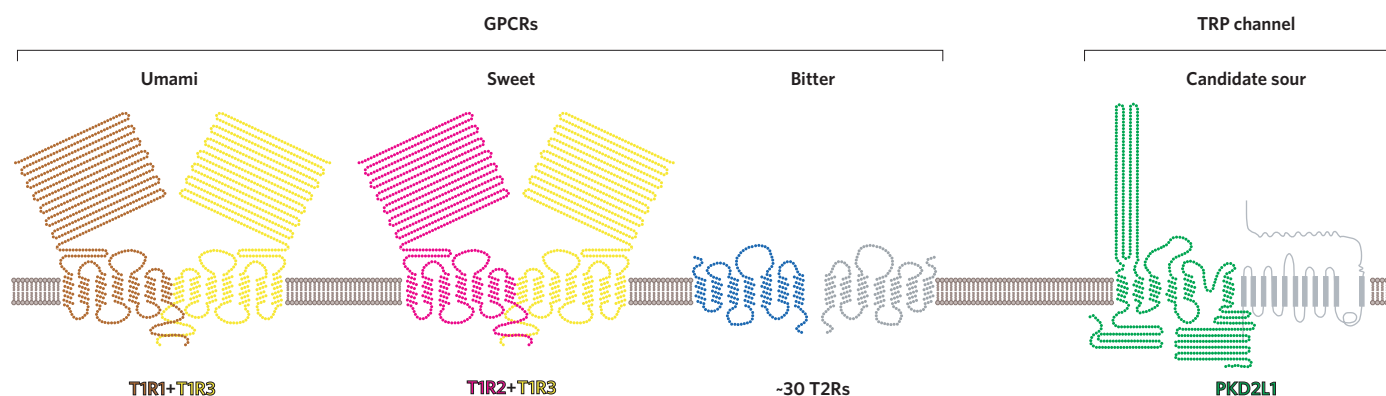


Figure 4 | Summary of receptors for umami, sweet, bitter and sour tastes. Schematic representation of taste receptors (and candidate receptors) mediating four of the five basic taste modalities. Although not indicated in the figure, responses to high concentrations of sugars, but not other sweet tastants, are also detected by T1R3 alone¹⁵. The grey T2R receptor is designed to illustrate the possibility that T2Rs, much like T1Rs, may function as heteromeric complexes. Similarly, the grey receptor next to PKD2L1 depicts a PKD1-family member as a candidate partner^{78–80}.

Table 1 | Tastant selectivity of candidate mammalian taste receptors

Tastant quality	Receptor(s)	Class of tastant	Examples of tastants
Umami	T1R1+T1R3	Amino acids	L-Glutamate, L-AP4, glycine*, L-amino acids*
		Nucleotide enhancers	IMP, GMP, AMP
Sweet	T1R2+T1R3	Sugars†	Sucrose, fructose, glucose, maltose
		Artificial sweeteners	Saccharin, acesulfame-K, cyclamate‡, aspartame‡
		D-amino acids	D-Phenylalanine, D-alanine, D-serine (also some selective L-amino acids)
		Sweet proteins‡	Monellin, thaumatin, curculin
Bitter§	T2R5¶		Cycloheximide
	T2R8¶, T2R4, T2R44		Denatonium
	T2R16		Salicin‡
	T2R38		PTC‡
	T2R43, T2R44		Saccharin
	Not known	Other toxic/noxious compounds	Quinine, strychnine, atropine
Sour	PKD2L1	Acids	Citric acid, tartaric acid, acetic acid, hydrochloric acid

*Preferentially activates mouse but not human receptors.

†High concentrations of sugars, but not other sweet tastants, can also be detected by T1R3 alone¹⁵.

‡Activates human but not mouse receptors and does not elicit behavioural responses in wild-type mice.

§About 30 T2Rs are involved in bitter-tastant recognition.

¶Mouse T2Rs; all others shown are human. There are 25 human and 35 mouse T2R bitter-taste receptors. For illustrative purposes we have included receptor–ligand matches for a number of de-orphaned T2Rs (for example, mouse T2R5 is the receptor for the protein synthesis inhibitor toxin cycloheximide).

Bitter

PLC- β 2 is required for sweet, umami and bitter tastes^{46,60}, so *Plc- β 2*-knockout animals are blind to stimuli from any of these three taste qualities (Fig. 3). If individual TRCs were tuned to a single taste quality, then restoring PLC function to a unique population of TRCs in *Plc*-knockout animals (for example T2R cells) should restore taste to a single taste modality (bitter taste in this example). By contrast, if these same cells were broadly tuned to sweet, amino-acid and bitter tastes, then restoring function to T2R cells (by expressing PLC) would restore taste to multiple modalities. Recently, Zhang and co-workers⁴⁶ showed that mice engineered to rescue PLC- β 2 function exclusively in T2R-expressing cells respond normally to bitter tastants but do not taste sweet or amino-acid stimuli. This 'selective rescue' experiment demonstrated both that activation of T2R cells is sufficient for normal bitter taste and that bitter taste is encoded independently of sweet and amino-acid tastes, with TRCs not broadly tuned across these modalities.

Bitter and sweet

To investigate the basis of sweet and bitter tastant recognition and coding, Zhao *et al.*¹⁵ and Mueller and colleagues⁴³ engineered mice that expressed a modified κ -opioid receptor (RASSL; receptor activated solely by a synthetic ligand⁸⁶) in either sweet or bitter cells. Animals expressing RASSL in sweet cells become selectively attracted to the synthetic-opioid-agonist spiradoline, a normally tasteless compound, demonstrating that activation of sweet-receptor-expressing cells, rather than the sweet receptor itself, results in the perception of sweetness. More importantly, these results showed unequivocally that activating a single cell type is sufficient to trigger specific taste responses. Does the same logic apply to bitter taste? Mueller and co-workers⁴³ tested this idea by generating mice in which the same RASSL receptor was targeted to bitter taste cells. Such mice showed marked aversion, rather than attraction, to spiradoline.

Together, these results demonstrated that a combinatorial pattern of activity (across-fibre pattern) is not needed to account for attraction or

aversion mediated by sweet- or bitter-sensing cells, thus strongly substantiating a labelled-line model of taste coding at the periphery. A final corollary to these findings is that expression of a sweet receptor in bitter cells should trigger behavioural aversion to sweet tastants, whereas expression of a bitter receptor in sweet cells should result in attraction to the bitter compound. Indeed, Mueller⁴³ engineered animals expressing a bitter receptor in sweet cells (Fig. 5) and these mice showed strong attraction to the cognate bitter compounds. Thus, the 'taste' of a sweet or bitter compound (in other words, the perception of sweet or bitter) is a reflection of the selective activation of T1R- versus T2R-expressing cells, rather than a property of the receptors or even of the tastant molecules.

Bitter, sweet and sour

Using a conceptually complementary approach to the functional rescue of subsets of TRCs described above, Huang and colleagues⁷⁸ eliminated entire populations of TRCs by genetically targeting expression of diphtheria toxin (DTA) to defined subsets of taste cells. Remarkably, animals expressing DTA in T1R2-, T2R- or PKD2L1-expressing cells showed extraordinarily specific taste deficits, with each exhibiting a marked loss of only a single taste quality (sweet, bitter and sour, respectively). Taken together, these studies reveal three fundamental features of taste coding at the periphery. First, they prove the functional segregation of individual taste modalities at the cellular level (as proposed in the original receptor expression studies^{8,46,78}). Second, they show the absolute requirement of T1R2-, T2R- and PKD2L1-cells for sweet, bitter and sour taste. Finally, they demonstrate that animals recognize and respond to taste cues (that is, encode and decode signals) without the need for combinatorial patterns of activity at the periphery (across-fibre models).

The exciting journey from detection to perception

The discovery that individual taste modalities are encoded by different TRCs should make it possible to mark the connectivity pathway for each taste quality individually. As a result, it should be possible to trace defined lines of information from the tongue to the cortex to understand not only where these signals go, but where and how they combine in the circuitry to choreograph taste and flavour.

Two recent reports have provided promising avenues to explore the connectivity between TRCs and central neuronal stations. In the

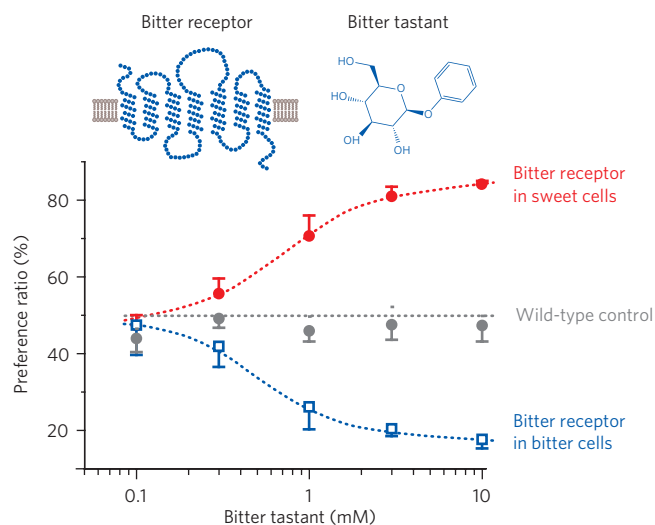


Figure 5 | Behavioural attraction and aversion are mediated by dedicated taste-receptor cells. Targeted expression of a novel bitter receptor to bitter (T2R-expressing) cells results in dose-dependent aversion to the specific bitter tastant (open blue squares). In marked contrast, directing expression of the same receptor to sweet cells produces animals that are strongly attracted to this bitter tastant (filled red circles). Control animals lacking the receptor (filled grey circles) are indifferent to the tastant.

first, Finger and colleagues⁸⁷ showed that all sweet, bitter, sour, salty and umami nerve responses are lost in *P2x2;P2x3* (purinergic receptor) double-knockout mice, implicating the purinergic agonist ATP as a potential neurotransmitter in taste⁸⁷. The availability of these taste-blind mice may now provide an experimental platform to engineer animals with function in defined sets of fibres, and therefore track the response of selective ganglion neurons. In the second, Sugita and Shiba⁸⁸ used a genetically encoded fluorescent transneuronal tracer to help reveal the circuitry linking TRCs to the brain⁸⁸. Interpretation of the results of this study was complicated by several technical difficulties, including poor transmission of the tracer, the use of a single label for both sweet and bitter pathways (thus necessitating comparison between animals), and the lack of anatomical correlates for many of the labelled neurons. However, this type of approach^{89,90} will be valuable in helping decipher the neural wiring for sweet, umami, bitter, sour and salty tastes.

Everyone appreciates that taste perception varies according to context. For example, the addition of sugar to lemon juice masks its sourness without affecting its acidity. More importantly, our perception of taste is significantly extended by other inputs — such as olfactory, visual and somatosensory — as well as prior experience, satiety and hunger⁹¹. This indicates that combination and comparison across taste qualities, together with information from other sensory systems, must ultimately converge to orchestrate the final percept in the brain. Although electrophysiological studies of the response profiles of brainstem, thalamic or cortical taste neurons are providing important insight into the basic properties of the central taste circuitry^{92,93}, the inherent technical difficulties in obtaining such data (often resulting in small sample sizes), have so far prevented the formulation of a true consensus view⁹⁴. We expect that molecular genetic and physiological approaches using novel reporters and genetically encoded activators and inhibitors of neuronal activity, combined with functional imaging at single-cell resolution^{95–100}, will be invaluable in helping to decipher how information flows from the tongue to sensory integration centres in the brain, ultimately, to dictate behaviour.

- Smith, D. V. & St John, S. J. Neural coding of gustatory information. *Curr. Opin. Neurobiol.* **9**, 427–435 (1999).
- Erickson, R. P., Covey, E. & Doetsch, G. S. Neuron and stimulus typologies in the rat gustatory system. *Brain Res.* **196**, 513–519 (1980).
- Erickson, R. P. The evolution of neural coding ideas in the chemical senses. *Physiol. Behav.* **69**, 3–13 (2000).
- Caicedo, A., Kim, K. N. & Roper, S. D. Individual mouse taste cells respond to multiple chemical stimuli. *J. Physiol. (Lond.)* **544**, 501–509 (2002).
- Smith, D. V., John, S. J. & Boughter, J. D. Neuronal cell types and taste quality coding. *Physiol. Behav.* **69**, 77–85 (2000).
- Hoon, M. A. et al. Putative mammalian taste receptors: a class of taste-specific GPCRs with distinct topographic selectivity. *Cell* **96**, 541–551 (1999).
- Bachmanov, A. A. et al. Positional cloning of the mouse saccharin preference (*Sac*) locus. *Chem. Senses* **26**, 925–933 (2001).
- Nelson, G. et al. Mammalian sweet taste receptors. *Cell* **106**, 381–390 (2001).
- Nelson, G. et al. An amino-acid taste receptor. *Nature* **416**, 199–202 (2002).
- Li, X. et al. Human receptors for sweet and umami taste. *Proc. Natl Acad. Sci. USA* **99**, 4692–4696 (2002).
- Kitagawa, M., Kusakabe, Y., Miura, H., Ninomiya, Y. & Hino, A. Molecular genetic identification of a candidate receptor gene for sweet taste. *Biochem. Biophys. Res. Commun.* **283**, 236–242 (2001).
- Max, M. et al. *Tas1r3*, encoding a new candidate taste receptor, is allelic to the sweet responsiveness locus *Sac*. *Nature Genet.* **28**, 58–63 (2001).
- Montmayeur, J. P., Liberles, S. D., Matsunami, H. & Buck, L. B. A candidate taste receptor gene near a sweet taste locus. *Nature Neurosci.* **4**, 492–498 (2001).
- Sainz, E., Korley, J. N., Battey, J. F. & Sullivan, S. L. Identification of a novel member of the T1R family of putative taste receptors. *J. Neurochem.* **77**, 896–903 (2001).
- Zhao, G. Q. et al. The receptors for mammalian sweet and umami taste. *Cell* **115**, 255–266 (2003).
- Pin, J. P. & Acher, F. The metabotropic glutamate receptors: structure, activation mechanism and pharmacology. *Curr. Drug Targets CNS Neurol. Disord.* **1**, 297–317 (2002).
- Kunishima, N. et al. Structural basis of glutamate recognition by a dimeric metabotropic glutamate receptor. *Nature* **407**, 971–977 (2000).
- Fuller, J. L. Single-locus control of saccharin preference in mice. *J. Hered.* **65**, 33–36 (1974).
- Lush, I. E. The genetics of tasting in mice. VI. Saccharin, acesulfame, dulcin and sucrose. *Genet. Res.* **53**, 95–99 (1989).
- Li, X. et al. High-resolution genetic mapping of the saccharin preference locus (*Sac*) and the putative sweet taste receptor (*TIR1*) gene (*Gpr70*) to mouse distal Chromosome 4. *Mamm. Genome* **12**, 13–16 (2001).
- Danilova, V., Hellekant, G., Tinti, J. M. & Nofre, C. Gustatory responses of the hamster *Mesocricetus auratus* to various compounds considered sweet by humans. *J. Neurophysiol.* **80**, 2102–2112 (1998).
- Xu, H. et al. Different functional roles of T1R subunits in the heteromeric taste receptors. *Proc. Natl Acad. Sci. USA* **101**, 14258–14263 (2004).
- Jiang, P. et al. Molecular mechanisms of sweet receptor function. *Chem. Senses* **30** (Suppl. 1), i17–i18 (2005).
- Jiang, P. et al. The cysteine-rich region of T1R3 determines responses to intensely sweet proteins. *J. Biol. Chem.* **279**, 45068–45075 (2004).
- Damak, S. et al. Detection of sweet and umami taste in the absence of taste receptor T1r3. *Science* **301**, 850–853 (2003).
- Li, X. et al. Pseudogenization of a sweet-receptor gene accounts for cats' indifference toward sugar. *PLoS Genet.* **1**, 27–35 (2005).
- Iwasaki, K., Kasahara, T. & Sato, M. Gustatory effectiveness of amino acids in mice: behavioral and neurophysiological studies. *Physiol. Behav.* **34**, 531–542 (1985).
- Iwasaki, K. & Sato, M. A. Taste preferences for amino acids in the house musk shrew, *Suncus murinus*. *Physiol. Behav.* **28**, 829–833 (1982).
- Pritchard, T. C. & Scott, T. R. Amino acids as taste stimuli. I. Neural and behavioral attributes. *Brain Res.* **253**, 81–92 (1982).
- Ikeda, K. New seasonings. *Chem. Senses* **27**, 847–849 (2002).
- Yamaguchi, S. The synergistic taste effect of monosodium glutamate and disodium 5'-inosinate. *J. Food Sci.* **32**, 473–478 (1967).
- Adler, E. et al. A novel family of mammalian taste receptors. *Cell* **100**, 693–702 (2000).
- Matsunami, H., Montmayeur, J. P. & Buck, L. B. A family of candidate taste receptors in human and mouse. *Nature* **404**, 601–604 (2000).
- Lush, I. E. & Holland, G. The genetics of tasting in mice. V. Glycine and cycloheximide. *Genet. Res.* **52**, 207–212 (1988).
- Reed, D. R. et al. Localization of a gene for bitter-taste perception to human chromosome 5p15. *Am. J. Hum. Genet.* **64**, 1478–1480 (1999).
- Chandrasekar, J. et al. T2Rs function as bitter taste receptors. *Cell* **100**, 703–711 (2000).
- Bufe, B., Hofmann, T., Krautwurst, D., Raguse, J. D. & Meyerhof, W. The human TAS2R16 receptor mediates bitter taste in response to β -glucopyranosides. *Nature Genet.* **32**, 397–401 (2002).
- Pronin, A. N., Tang, H., Connor, J. & Keung, W. Identification of ligands for two human bitter T2R receptors. *Chem. Senses* **29**, 583–593 (2004).
- Kuhn, C. et al. Bitter taste receptors for saccharin and acesulfame K. *J. Neurosci.* **24**, 10260–10265 (2004).
- Behrens, M. et al. The human taste receptor hTAS2R14 responds to a variety of different bitter compounds. *Biochem. Biophys. Res. Commun.* **319**, 479–485 (2004).
- Wooding, S. et al. Independent evolution of bitter-taste sensitivity in humans and chimpanzees. *Nature* **440**, 930–934 (2006).
- Kim, U. K. et al. Positional cloning of the human quantitative trait locus underlying taste sensitivity to phenylthiocarbamide. *Science* **299**, 1221–1225 (2003).
- Mueller, K. L. et al. The receptors and coding logic for bitter taste. *Nature* **434**, 225–229 (2005).
- Shi, P. & Zhang, J. Contrasting modes of evolution between vertebrate sweet/umami receptor genes and bitter receptor genes. *Mol. Biol. Evol.* **23**, 292–300 (2006).
- Go, Y., Satta, Y., Takenaka, O. & Takahata, N. Lineage-specific loss of function of bitter taste receptor genes in humans and nonhuman primates. *Genetics* **170**, 313–326 (2005).
- Zhang, Y. et al. Coding of sweet, bitter, and umami tastes: different receptor cells sharing similar signaling pathways. *Cell* **112**, 293–301 (2003).
- Spector, A. C. & Kopka, S. L. Rats fail to discriminate quinine from denatonium: implications for the neural coding of bitter-tasting compounds. *J. Neurosci.* **22**, 1937–1941 (2002).
- Chaudhari, N. & Roper, S. D. Molecular and physiological evidence for glutamate (umami) taste transduction via a G protein-coupled receptor. *Ann. NY Acad. Sci.* **855**, 398–406 (1998).
- Kinnamon, S. C. A plethora of taste receptors. *Neuron* **25**, 507–510 (2000).
- Smith, D. V. & Margolskee, R. F. Making sense of taste. *Sci. Am.* **284**, 32–39 (2001).
- Brunet, L. J., Gold, G. H. & Ngai, J. General anosmia caused by a targeted disruption of the mouse olfactory cyclic nucleotide-gated cation channel. *Neuron* **17**, 681–693 (1996).
- Damak, S. et al. Trpm5 null mice respond to bitter, sweet, and umami compounds. *Chem. Senses* **31**, 253–264 (2006).
- McLaughlin, S. K., McKinnon, P. J. & Margolskee, R. F. Gustducin is a taste-cell-specific G protein closely related to the transducins. *Nature* **357**, 563–569 (1992).
- Kusakabe, Y. et al. Comprehensive study on G protein α -subunits in taste bud cells, with special reference to the occurrence of Gai2 as a major G α species. *Chem. Senses* **25**, 525–531 (2000).
- Huang, L. et al. G13 colocalizes with gustducin in taste receptor cells and mediates IP3 responses to bitter denatonium. *Nature Neurosci.* **2**, 1055–1062 (1999).
- Rossler, P., Kroner, C., Freitag, J., Noe, J. & Breer, H. Identification of a phospholipase C β subtype in rat taste cells. *Eur. J. Cell Biol.* **77**, 253–261 (1998).
- Perez, C. A. et al. A transient receptor potential channel expressed in taste receptor cells. *Nature Neurosci.* **5**, 1169–1176 (2002).
- Wong, G. T., Gannon, K. S. & Margolskee, R. F. Transduction of bitter and sweet taste by gustducin. *Nature* **381**, 796–800 (1996).
- Ruiz, C. J., Wray, K., Delay, E., Margolskee, R. F. & Kinnamon, S. C. Behavioral evidence for a role of α -gustducin in glutamate taste. *Chem. Senses* **28**, 573–579 (2003).
- Dotson, C. D., Roper, S. D. & Spector, A. C. PLC β 2-independent behavioral avoidance of prototypical bitter-tasting ligands. *Chem. Senses* **30**, 593–600 (2005).
- Varkevisser, B. & Kinnamon, S. C. Sweet taste transduction in hamster: role of protein kinases. *J. Neurophysiol.* **83**, 2526–2532 (2000).
- Rosenzweig, S., Yan, W., Dasso, M. & Spielman, A. I. Possible novel mechanism for bitter taste mediated through cGMP. *J. Neurophysiol.* **81**, 1661–1665 (1999).
- Bernhardt, S. J., Naim, M., Zehavi, U. & Lindemann, B. Changes in IP3 and cytosolic Ca²⁺ in response to sugars and non-sugar sweeteners in transduction of sweet taste in the rat. *J. Physiol. (Lond.)* **490**, 325–336 (1996).
- Striemi, B. J., Pace, U., Zehavi, U., Naim, M. & Lancet, D. Sweet tastants stimulate adenylate cyclase coupled to GTP-binding protein in rat tongue membranes. *Biochem. J.* **260**, 121–126 (1989).
- Gilbertson, T. A. & Boughter, J. D. Taste transduction: appetizing times in gustation. *Neuroreport* **14**, 905–911 (2003).

66. Avenet, P., Hofmann, F. & Lindemann, B. Transduction in taste receptor cells requires cAMP-dependent protein kinase. *Nature* **331**, 351–354 (1988).
67. Talavera, K. *et al.* Heat activation of TRPM5 underlies thermal sensitivity of sweet taste. *Nature* **438**, 1022–1025 (2005).
68. Heck, G. L., Mierson, S. & DeSimone, J. A. Salt taste transduction occurs through an amiloride-sensitive sodium transport pathway. *Science* **223**, 403–405 (1984).
69. Avenet, P. & Lindemann, B. Amiloride-blockable sodium currents in isolated taste receptor cells. *J. Membr. Biol.* **105**, 245–255 (1988).
70. Lyall, V. *et al.* The mammalian amiloride-insensitive non-specific salt taste receptor is a vanilloid receptor-1 variant. *J. Physiol. (Lond.)* **558**, 147–159 (2004).
71. Stevens, D. R. *et al.* Hyperpolarization-activated channels HCN1 and HCN4 mediate responses to sour stimuli. *Nature* **413**, 631–635 (2001).
72. Ugawa, S. *et al.* Receptor that leaves a sour taste in the mouth. *Nature* **395**, 555–556 (1998).
73. Lin, W., Burks, C. A., Hansen, D. R., Kinnamon, S. C. & Gilbertson, T. A. Taste receptor cells express pH-sensitive leak K⁺ channels. *J. Neurophysiol.* **92**, 2909–2919 (2004).
74. Richter, T. A., Dvoryanchikov, G. A., Chaudhari, N. & Roper, S. D. Acid-sensitive two-pore domain potassium (K2P) channels in mouse taste buds. *J. Neurophysiol.* **92**, 1928–1936 (2004).
75. Waldmann, R., Champigny, G., Bassilana, F., Heurteaux, C. & Lazdunski, M. A proton-gated cation channel involved in acid-sensing. *Nature* **386**, 173–177 (1997).
76. Lyall, V. *et al.* Basolateral Na⁺–H⁺ exchanger-1 in rat taste receptor cells is involved in neural adaptation to acidic stimuli. *J. Physiol. (Lond.)* **556**, 159–173 (2004).
77. Cummings, T. A. & Kinnamon, S. C. Apical K⁺ channels in *Necturus* taste cells. Modulation by intracellular factors and taste stimuli. *J. Gen. Physiol.* **99**, 591–613 (1992).
78. Huang, A. L. *et al.* The cells and logic for mammalian sour taste detection. *Nature* **442**, 934–938 (2006).
79. Lopezjimenez, N. D. *et al.* Two members of the TRPP family of ion channels, Pkd1l3 and Pkd2l1, are co-expressed in a subset of taste receptor cells. *J. Neurochem.* **98**, 68–77 (2006).
80. Ishimaru, Y. *et al.* Transient receptor potential family members PKD1L3 and PKD2L1 form a candidate sour taste receptor. *Proc. Natl Acad. Sci. USA* **103**, 12569–12574 (2006).
81. Lahiri, S. & Forster, R. E. CO₂/H⁺ sensing: peripheral and central chemoreception. *Int. J. Biochem. Cell Biol.* **35**, 1413–1435 (2003).
82. Vigh, B. *et al.* The system of cerebrospinal fluid-contacting neurons. Its supposed role in the nonsynaptic signal transmission of the brain. *Histol. Histopathol.* **19**, 607–628 (2004).
83. Gilbertson, T. A., Boughter, J. D., Zhang, H. & Smith, D. V. Distribution of gustatory sensitivities in rat taste cells: whole-cell responses to apical chemical stimulation. *J. Neurosci.* **21**, 4931–4941 (2001).
84. Sato, T. & Beidler, L. M. Broad tuning of rat taste cells for four basic taste stimuli. *Chem. Senses* **22**, 287–293 (1997).
85. Richter, T. A., Caicedo, A. & Roper, S. D. Sour taste stimuli evoke Ca²⁺ and pH responses in mouse taste cells. *J. Physiol. (Lond.)* **547**, 475–483 (2003).
86. Redfern, C. H. *et al.* Conditional expression and signaling of a specifically designed Gi-coupled receptor in transgenic mice. *Nature Biotechnol.* **17**, 165–169 (1999).
87. Finger, T. E. *et al.* ATP signaling is crucial for communication from taste buds to gustatory nerves. *Science* **310**, 1495–1499 (2005).
88. Sugita, M. & Shiba, Y. Genetic tracing shows segregation of taste neuronal circuitries for bitter and sweet. *Science* **309**, 781–785 (2005).
89. Zou, Z., Horowitz, L. F., Montmayeur, J. P., Snapper, S. & Buck, L. B. Genetic tracing reveals a stereotyped sensory map in the olfactory cortex. *Nature* **414**, 173–179 (2001).
90. Kuze, B., Matsuyama, K., Matsui, T., Miyata, H. & Mori, S. Segment-specific branching patterns of single vestibulospinal tract axons arising from the lateral vestibular nucleus in the cat: A PHA-L tracing study. *J. Comp. Neurol.* **414**, 80–96 (1999).
91. Rolls, E. T. *The Brain and Emotion* (Oxford Univ. Press, USA, 2000).
92. Katz, D. B., Simon, S. A. & Nicolelis, M. A. Dynamic and multimodal responses of gustatory cortical neurons in awake rats. *J. Neurosci.* **21**, 4478–4489 (2001).
93. Di Lorenzo, P. M. The neural code for taste in the brain stem: response profiles. *Physiol. Behav.* **69**, 87–96 (2000).
94. Spector, A. C. & Travers, S. P. The representation of taste quality in the mammalian nervous system. *Behav. Cogn. Neurosci. Rev.* **4**, 143–191 (2005).
95. Zhang, J., Campbell, R. E., Ting, A. Y. & Tsien, R. Y. Creating new fluorescent probes for cell biology. *Nature Rev. Mol. Cell Biol.* **3**, 906–918 (2002).
96. Choi, G. B. *et al.* Lhx6 delineates a pathway mediating innate reproductive behaviors from the amygdala to the hypothalamus. *Neuron* **46**, 647–660 (2005).
97. Miesenböck, G. & Kevrekidis, I. G. Optical imaging and control of genetically designated neurons in functioning circuits. *Annu. Rev. Neurosci.* **28**, 533–563 (2005).
98. Gosgnach, S. *et al.* V1 spinal neurons regulate the speed of vertebrate locomotor outputs. *Nature* **440**, 215–219 (2006).
99. Gogos, J. A., Osborne, J., Nemes, A., Mendelsohn, M. & Axel, R. Genetic ablation and restoration of the olfactory topographic map. *Cell* **103**, 609–620 (2000).
100. Brecht, M. *et al.* Novel approaches to monitor and manipulate single neurons *in vivo*. *J. Neurosci.* **24**, 9223–9227 (2004).

Acknowledgements We thank a group of extraordinary students, postdoctoral fellows and research technicians in our laboratories, who joined us on this wonderful journey of mammalian taste research beginning in the fall of 1997. N.J.P.R. is an investigator in the Intramural program at the NIH, NIDCR. C.S.Z. is an investigator of the Howard Hughes Medical Institute.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence should be addressed to C.S.Z. or N.J.P.R. (czuker@ucsd.edu; nick.ryba@nih.gov).

Comparative chemosensation from receptors to ecology

Cornelia I. Bargmann¹

Odour perception is initiated by specific interactions between odorants and a large repertoire of receptors in olfactory neurons. During the past few years, considerable progress has been made in tracing olfactory perception from the odorant receptor protein to the activity of olfactory neurons to higher processing centres and, ultimately, to behaviour. The most complete picture is emerging for the simplest olfactory system studied — that of the fruitfly *Drosophila melanogaster*. Comparison of rodent, insect and nematode olfaction reveals surprising differences and unexpected similarities among chemosensory systems.

All living organisms detect chemicals in their environments. Animal olfactory systems are impressively complex, and can detect almost all airborne molecules and distinguish between them with great precision. The odorant receptor genes that permit chemosensation were discovered in 1991 through insightful use of molecular biology¹. Fifteen years later, the process of receptor-gene discovery continues in a new form: because the large gene families involved in odour detection are highly diverse in different animals, new receptor genes are identified daily by genome sequencing projects.

Olfaction begins with the detection of odours by G-protein-coupled receptors (GPCRs), which are encoded by up to 5–10% of an animal's genes. Biochemistry, genetics and physiology have been used to probe the functions of a few of these many receptors, giving a first view of the relationship between odours and their molecular detectors. Although odour recognition is sometimes presented as a chemical problem of the carbon-chain lengths and functional groups the odour contains, it is fundamentally a biochemical process in which a complete chemical is matched with potential receptors. Odours are beginning to be understood as ligands for these receptors, and not just as chemical structures.

The matching of odours with receptors results in activation of olfactory receptor neurons. In mammals and insects, each of these expresses only one or a few receptor genes. These neurons convey odour information to second-order neurons, in which information is integrated and refined. The mapping of odours onto the full array of receptor proteins, receptor neurons and second-order neurons is best understood in the simple olfactory system of *Drosophila*. Because olfactory research in *Drosophila* has progressed rapidly in recent years, it is a focus of this review.

Higher-order mapping in the brain couples odours to innate or learned behaviours. The olfactory system can sense an astonishing array of chemicals, but most odours are produced by living organisms, and each animal specializes in detecting biologically relevant features of its own environment. A view of olfaction that incorporates evolutionary constraints explains non-intuitive features of olfactory recognition and processing.

Chemosensory receptor proteins are diverse

Mammalian olfactory receptors are G-protein-coupled receptors

Olfaction in the mammalian olfactory epithelium is mediated by a large family of GPCRs from the rhodopsin-related receptor superfamily. These are known collectively as odorant receptors (Ors)¹ (Fig. 1a). Receptors from two additional GPCR families mediate olfaction

in the specialized vomeronasal organ (see page 308); and two smaller GPCR families mediate sweet and bitter taste (see page 288). The total number of odorant receptor genes varies widely across species; mice have about 900 functional odorant receptor genes (and ~1,200 chemoreceptor genes in all), whereas humans have about 350 functional odorant receptor genes and have lost the vomeronasal receptors during evolution². These differences result both from duplication of genes in the mouse lineage and from extensive loss of genes in the human lineage³. Odorant receptor genes also diversify quickly in comparison with other genes, so it can be challenging to recognize orthologous gene pairs even between humans and mice². The odorant receptor family exists from fish through to mammals, but is not present in the simple chordate ancestor *Ciona*.

Invertebrate olfactory receptors have distinct origins

Like other gene families, odorant receptor families have expanded over time by gene duplication and divergence. In addition, however, new odorant receptor families have been recruited several times during evolution (Fig. 1b). Invertebrate olfaction is mediated by receptor families that are evolutionarily distinct from vertebrate receptors. The genome of the nematode worm *Caenorhabditis elegans* encodes more than 1,500 predicted GPCRs in nematode-specific families of the rhodopsin-related GPCR superfamily⁴. These genes subserve both taste and smell, and perhaps other functions; most are expressed in chemosensory neurons, but ~20% are expressed in other tissues^{5–8}. There may be more than 1,000 functional chemosensory receptors in nematodes, a surprisingly large number for a small animal. *C. elegans* also uses non-GPCRs for chemosensation — it senses environmental oxygen levels using soluble guanylate cyclase homologues that bind directly to molecular oxygen through a haem group^{9,10}. Receptor-like guanylate cyclases may also function as nematode chemoreceptors¹¹. The diversity of nematode receptors highlights the opportunistic nature of chemosensation, and its ability to recruit many receptor types to a common task.

The genome of the fruitfly *Drosophila melanogaster* has only 62 odorant receptors encoded by 60 genes, and 68 gustatory receptors (GRs) encoded by another 60 genes^{12,13}. The malaria-carrying mosquito *Anopheles* has a similar molecular complement of 85 odorant receptors and 76 gustatory receptors, suggesting that small numbers of receptors may be the norm in insects (see page 302). The insect receptor genes

¹Howard Hughes Medical Institute, The Rockefeller University, 1230 York Avenue, New York, New York 10021, USA.

are molecularly unrelated to any other known gene, and seem to be of different evolutionary origin from canonical GPCRs (Fig. 1b). Indeed, it is questionable whether these molecules are GPCRs at all. Antibody-tagging experiments suggest that the membrane topology of at least two of the *Drosophila* odorant receptors is altered with respect to classical GPCRs, which have extracellular amino termini — the N termini of fly odorant receptors are cytoplasmic¹⁴ (Fig. 1a). This experimental evidence is bolstered by computational analysis of GPCR sequences — hidden Markov models suggest that *Drosophila* odorant receptor membrane topology differs from that of known GPCRs¹⁵. Expression of *Drosophila* odorant receptors in heterologous cells can result in signaling pathway activation¹⁶; whether this is a G-protein-mediated action of the odorant receptor is an interesting question.

The small complement of *Drosophila* odorant receptors simplifies the mapping of odorant specificity by an order of magnitude compared with mice or *C. elegans*. The near-complete set of *Drosophila* odorant receptors has been analysed at the level of peripheral expression, functional response and anatomical mapping to the antennal lobe. This is unlikely to be equalled in other animals in the near future.

Odour-receptor interactions in the olfactory system

A GPCR-based view of odour-receptor recognition

A few dozen mammalian odorant receptors have known odour ligands^{17–24}. With respect to ligand binding, odorant receptor properties are consistent with those of rhodopsin and related GPCRs that have been the subject of extensive biochemical and structural analysis^{25,26}. Rhodopsin-like GPCRs exist in one of two main conformations: an

inactive conformation, and an active conformation that interacts productively with an intracellular heterotrimeric G protein. The transition between these conformations occurs through the movement of various membrane-spanning domains. This movement is nucleated around a small molecule or peptide that interacts with several domains simultaneously^{25,26}. Although a subset of membrane-spanning regions are most frequently associated with ligand binding, different agonists of the same GPCR need not bind to the same site within the receptor²⁷. GPCR agonists stabilize the active form of the receptor, whereas antagonists can block agonist binding and inverse agonists stabilize the inactive conformation. Partial GPCR agonists can stabilize subconformations of the active state, allowing different ligands to have different effects. Thus the efficacy of a GPCR agonist does not depend on a single functional group or feature of the chemical, or even on its affinity for receptor binding, but rather on its ability to stabilize the active functional GPCR state and destabilize the inactive state.

The properties of odorant receptors are likely to follow the rules described above for rhodopsin and related GPCRs. For example, the functional properties of the known odorant receptor ligands are distributed throughout the entire structure of the chemical, not in individual features such as functional groups, and many regions of the receptor contribute to binding²⁸. Flexible long-chain aliphatic molecules can adopt many conformations, so odorant receptors that sense aliphatic molecules can also recognize families of related molecules¹⁹. Rigid ring structures can mediate more selective odorant receptor activation²⁰. In addition, specific antagonists can block receptor activation by odour agonists^{23,24,29}.

Broadly and narrowly tuned receptors contribute to olfaction

In *Drosophila*, the properties of 31 classes of olfactory neurons have been recorded directly by electrophysiology or calcium imaging^{30–34}. In addition, 35 of the 62 odorant receptor genes have been systematically expressed in a single ‘host’ class of olfactory neuron and characterized by extracellular recording after exposure to up to 110 odorants^{35–37}. This work shows that fruitflies detect many odorants with a small number of odorant receptors and a combination of strategies. Some such receptors are highly selective: Or82a responded strongly to only one of 110 chemicals and weakly to five additional chemicals. Others were broadly activated by up to 30% of the odorants presented, including odorants with little structural similarity. Some odorant receptors were found have intermediate properties. Fruit odours activate about two-thirds of the receptors, highlighting the fly’s ecological speciality. Odour-independent basal activity is prominent, and the basal activity of receptors can be inhibited by up to 30% of odorants³⁷. Different odours can generate distinct temporal responses from the same receptor, perhaps by acting as partial agonists.

Thus, the first step of *Drosophila* odour perception incorporates activation and inhibition, as well as generalist, intermediate and specialist receptors. Comparisons among the broadly tuned receptors that recognize aliphatic fruit odours may be important for food identification, and more specific information may be provided by the highly tuned receptors.

Little is known about the recognition specificity of *C. elegans* chemoreceptors. Forward genetic screens for olfactory mutants identified one GPCR, ODR-10, as the receptor for the volatile odorant diacetyl³⁸. *odr-10* mutants have a 100-fold reduced sensitivity to diacetyl but are sensitive to all other tested odorants³⁸. As mentioned above, soluble guanylate cyclases seem to function as oxygen sensors^{9,10}. These results suggest that at least some of the many *C. elegans* receptors are narrowly tuned to specific chemicals.

Expression strategies for chemosensory receptors

Like the receptor genes themselves, expression strategies for olfactory receptor genes differ among animals. The 10 million vertebrate olfactory neurons follow an apparently strict expression pattern of one odorant receptor gene being expressed per cell^{39,40} (Fig. 2a). Transcriptional elements limit receptor expression to a broad zone of the olfactory epithelium but do not explain the precise single-cell pattern^{41–43}. After one

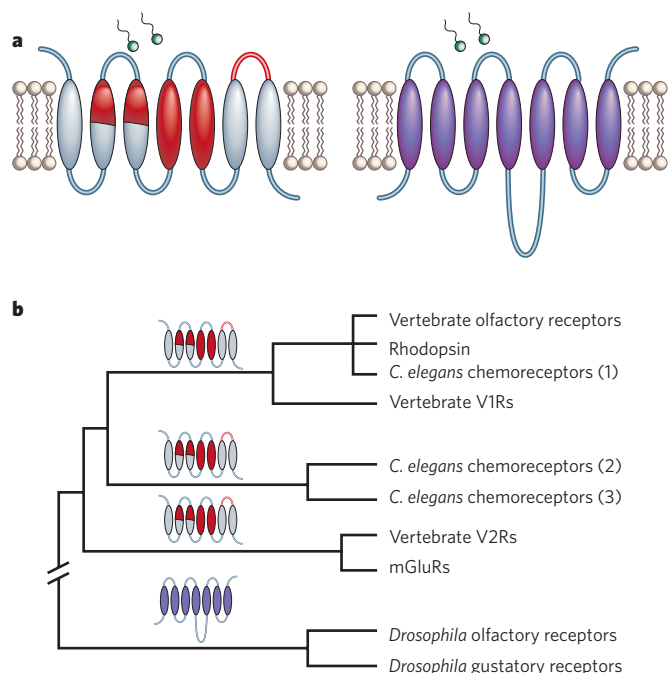


Figure 1 | Olfactory receptor diversity. **a**, Vertebrate olfactory receptors and *C. elegans* chemoreceptors are classical GPCRs with extracellular N termini (left), whereas *Drosophila* receptors have an inverted topology (right). Red regions denote regions of the vertebrate olfactory receptors that are highly diverse between different receptors, suggesting a role in ligand recognition. Small odour molecules reach the receptor from the extracellular environment (top). (Data from ref. 15.) **b**, Vertebrate olfactory receptors and some nematode chemoreceptors derive from the rhodopsin family of GPCRs; other nematode chemoreceptors and vertebrate vomeronasal V1R and V2R receptors derive from more distant branches of the GPCR tree. *Drosophila* gustatory and olfactory receptors are distant from GPCRs in evolutionary origin, and are possibly more closely related to 6-transmembrane-domain ion channels. mGluR, metabotropic glutamate receptor. 1, 2 and 3 represent different families of putative chemoreceptor genes.

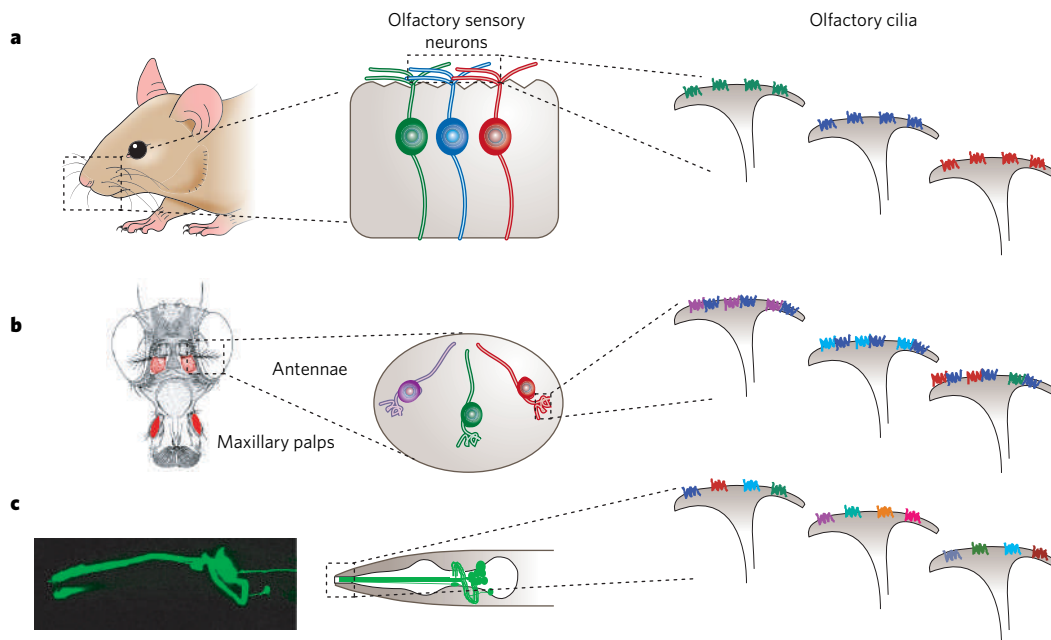


Figure 2 | Different receptor expression strategies in different animals. The locations of the main olfactory tissues in the mouse (a), *Drosophila* (b) and *C. elegans* (c) are shown. The odorant detection machinery is localized to the olfactory cilia, which are specializations of the olfactory neurons that come into direct contact with the environment (centre). The cilia of three representative olfactory neurons in each species are illustrated in the right-hand panel. Colours denote different odorant receptor proteins in the cilia. **a**, In mammals, olfactory neurons reside in the nasal epithelium. Each olfactory neuron expresses one type of odorant receptor, a GPCR. **b**, In *Drosophila*, olfactory neurons are localized to antennae and maxillary palps. Most olfactory neurons express one specific olfactory receptor and one ubiquitous one, Or83b (dark blue). Occasionally, neurons express two specific olfactory receptors in addition to Or83b (far right). The orientation of fly olfactory receptors is inverted with respect to classical GPCRs (Fig. 1a). (Fly head image modified, with permission, from Kyotofly Kit, Kyoto Institute of Technology, Japan.) **c**, In *C. elegans*, each olfactory neuron expresses several different GPCRs. The cell body is in the head and the cilia are at the tip of the nose. (Image courtesy of Jason Kennerdell, Rockefeller University, New York, USA.)

functional receptor is chosen, it seems to prevent expression of other receptor genes through feedback mechanisms^{40,44,45}.

Most of the ~2,600 *Drosophila* olfactory neurons express two receptor genes — one specific to the cell type, and a second, ubiquitous gene known as *Or83b*, whose protein product dimerizes with the specific receptor and mediates its transport to olfactory cilia^{13,16,46} (Fig. 2b). This arrangement can be thought of as functionally equivalent to the one receptor gene per cell pattern of vertebrates, but *Drosophila* allows more flexibility in expression patterns. For example, six to eight classes of olfactory neuron associated with the antennae express two odorant receptor genes in addition to *Or83b*^{32,47–50}. Unlike the stochastic vertebrate odorant receptor gene-expression pattern, expression in *Drosophila* is established by a deterministic developmental pathway, so that neurons in a particular sensillum always express the same receptor gene. Feedback is unlikely to have a role in *Drosophila* receptor expression, as transgenes can be used to misexpress an odorant receptor gene in a cell containing a second, endogenous one^{32,49}.

A radically different expression pattern is seen in *C. elegans*, which has hundreds of chemoreceptor genes but only 32 chemosensory neurons. In *C. elegans*, a single neuron expresses many different chemoreceptor genes (Fig. 2c); for example, at least nine different genes are strongly expressed in ADL neurons^{5,7}. Most chemoreceptor genes are strongly expressed in a single left–right pair of chemosensory neurons, although some receptors are expressed in only one cell or in many cell types^{5–8,11}. These complex gene-expression patterns are established mainly through cell-type-specific transcription factors and micro RNAs^{51–53}. One case of stochastic receptor expression is known⁵⁴.

Despite the different gene-expression strategies, a common theme in all these animals is the great diversity of receptor cell types. Further diversity comes from regulated odorant receptor expression. The expression of certain chemoreceptor genes in *C. elegans* is regulated by pheromones and food, allowing the animal to adjust its chemical sensitivity to the environment^{55,56}. In the malaria mosquito, the expression of certain receptor genes is downregulated after a blood meal⁵⁷, suggesting

that insects, like nematodes, may use gene-expression changes at the periphery to alter behavioural sensitivity to odours. *Drosophila* larvae express odorant receptor genes that are largely non-overlapping with those expressed in adults^{36,58}, and fish odorant receptors are also induced at different times during development⁵⁹. Regulated receptor expression has not yet been described in mammals.

Second-order coding of olfactory information

Convergence of sensory inputs onto glomeruli

A common feature of olfactory coding in *Drosophila* and vertebrates is the convergence of olfactory neurons expressing the same receptor to glomeruli in the second processing station, the antennal lobe or olfactory bulb (Fig. 3). In glomeruli, neurons expressing the same receptor converge on target neurons known as projection neurons in flies and as mitral or tufted cells in mice. Mapping of olfactory neurons to glomeruli is precise and invariant in flies; in mammals, the map is fairly precise but has small variations from one individual to another^{13,60–62}. The development of the mouse glomerular map is controlled by the odorant receptors themselves⁶³. The nematode *C. elegans* is unusual among animals in lacking a glomerular structure in its olfactory processing centres⁶⁴.

Although convergence on glomeruli is a common feature of olfaction, the mapping of odour sensation across the glomerular sheet varies in different animals. A complete map of the odorant receptor projection patterns to the antennal lobe has been generated for *Drosophila*, so the olfactory receptor inputs to each of the 43 glomeruli are known^{13,47,48}. There is no clear chemotopic map of odour features or structures in the fly antennal lobe; almost any odour can be detected by neurons projecting to almost any glomerulus^{37,47,48}. By contrast, in the mammalian olfactory bulb, adjacent glomeruli often recognize similar odours⁶⁵. The most extensive studies in mammals have been performed using intrinsic signal imaging, a method for mapping an animal's response to dozens of odours across about 200 glomeruli⁶⁶. The intrinsic signal reports presynaptic activity, and therefore the sensitivity of peripheral olfactory receptors⁶⁷. Intrinsic signal imaging has revealed clusters of glomeruli with

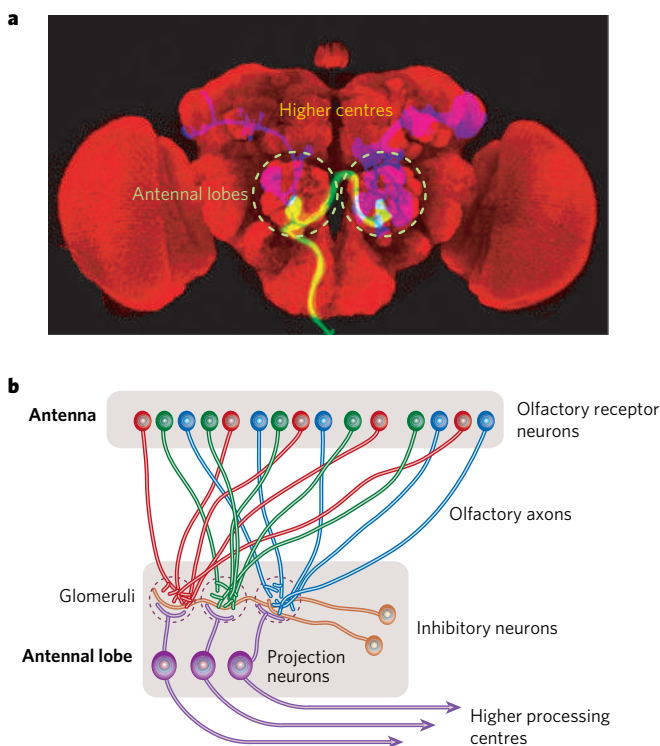


Figure 3 | Transition from the first to the second stage of olfactory processing. **a**, Superimposed images of antibody-stained *Drosophila* brains. In the *Drosophila* brain, olfactory receptor axons project from sensory organs to the antennal lobes, where they connect to projection neurons. A few olfactory receptor axons are labelled in yellow. Projection neurons (purple) in the antennal lobes (dashed green circles) project to higher brain centres, including the mushroom bodies and the lateral protocerebrum. (Image courtesy of Liquan Luo, Stanford University.) **b**, Schematic illustration of the glomerular organization of the *Drosophila* antennal lobe. The axons of neurons expressing the same type of odorant receptor converge on glomeruli in the antennal lobe. Blue, green and red are used to denote olfactory neurons expressing a common odorant receptor. Glomeruli are outlined in purple. In *Drosophila*, each excitatory projection neuron sends its dendrites to a single glomerulus, where it receives extensive input from a single class of olfactory receptor neuron, and sends an axon to higher brain centres (purple). The inhibitory (lateral) neurons (orange) in *Drosophila* project to multiple glomeruli. A similar organization of olfactory receptor axons and mitral cells (which are analogous to projection neurons) is found in the vertebrate olfactory bulb, the first processing station after the nose. Vertebrate interneurons, however, are more numerous and complex.

similar properties, and in some cases has shown a graded mapping of odour features across groups of adjacent glomeruli⁶⁵. A different pattern seems to exist in zebrafish, in which the olfactory bulb is divided into two distinct regions, one with large, well-separated glomeruli, and one with a clustered plexus of glomeruli that sense amino-acid odorants^{68,69}.

The transformation between sensory and second-order neurons

In flies and mice, a single projection neuron or mitral cell receives most of its sensory input from a single glomerulus, suggesting that the projection neuron or mitral cell will be activated primarily by the odorants recognized by one type of receptor^{13,47,48,60,61}. This can be tested in the fly olfactory system, because the odour specificity of the peripheral neurons is known and the odour specificity of defined projection neurons can be determined electrophysiologically. One such study has been conducted using the DM2 glomerulus, which is innervated by Or22a, one of the most broadly tuned odorant receptors. The projection neuron response in DM2 approximately matched the Or22a input for about half of the tested odorants⁷⁰. For most odorants, the projection neuron response was amplified and accelerated compared with the peripheral response,

a result expected on the basis of the ability of projection neurons to extract accurate information from correlated sensory inputs. Three odorants seemed to activate the projection neuron strongly without activating Or22a: octan-3-ol, phenyl acetaldehyde and butan-1-ol. However, in another study butan-1-ol was reported to activate Or22a — this mismatch may be the result of variations in sensitivity³⁷. The responses to phenyl acetaldehyde and octan-3-ol suggest that the projection neuron can respond to odours not sensed by the odorant receptor. The only odorant receptor known to sense phenyl acetaldehyde projects to the glomerulus adjacent to DM2, providing a possible nearby source of the novel odour information.

Less complete comparisons can be made for three other glomeruli for which the peripheral odorant receptors and projection neurons have been characterized by different groups. In two cases (VA3/Or67b and VM3/Or9a), the projection neuron was well matched to the odorant receptor for all tested odorants; in the third case (DM5/Or85a), the projection neuron response was very different from the response of the input neuron^{36,37,70,71}. Together, these results suggest that the projection neuron can function as a site of integration and transformation, not just as a collector of peripheral inputs.

In apparent contrast to the electrophysiology, calcium imaging studies of olfactory terminals and projection neuron dendrites seem to show a close match between the sensory input and the projection neuron output^{32,72} (Fig. 4). Discrepancies between the results of calcium imaging and electrophysiological studies can probably be partly explained by the bias of calcium imaging towards strong excitatory responses, which are likely to represent the primary odour quality but would not reflect subtle shaping of the odour response. Another contributory factor might be the differences between the olfactory dendrite and its axon terminals; odorant receptor axon terminals might be modulated by other inputs in the antennal lobe, and there are notable differences between reported dendrite responses and axon terminal responses for some olfactory receptor neurons^{32,37}.

In the antennal lobe, olfactory information is shaped by inhibitory γ -aminobutyric acid (GABA)-mediated interneurons of the antennal lobe known as local neurons. Local neurons modulate projection neuron properties and might also affect presynaptic olfactory receptor neuron terminals. Most inhibitory local neurons in *Drosophila* project widely, but not ubiquitously, across the antennal lobe, suggesting that they have a broad modulatory role⁷¹. Projection neurons have a substantial basal firing rate, and therefore the potential to be regulated by inhibitory inputs. Inhibitory episodes are prominent features in the response of projection neurons to odours, and much of this inhibition is blocked by a GABA type B antagonist that disrupts slow local-neuron signalling⁷¹. Complex local computations in and between glomeruli are also important contributors to mammalian olfaction, and these are the subject of ongoing research.

The tentative conclusion that can be drawn from this initial analysis is that the antennal lobe uses various processing strategies. Some odorant receptors and projection neurons are broadly tuned, whereas others are narrowly tuned. Projection neurons report many sensory inputs fairly faithfully, with an enhanced signal-to-noise ratio due to convergence and inhibition, and, with the assistance of local neurons, can transform sensory information to generate new messages. A subset of odorants activate projection neurons at frequencies of more than 100 Hz within milliseconds of odour presentation; most of these odorants strongly activate the input odorant receptor. Many odorants cause more subtle shifts in projection neuron behaviour, and a large number can inhibit the projection neuron for at least part of the time after presentation or removal of an odour; these are more likely to represent effects of antennal lobe processing.

Other insects — locusts, honeybees and moths — have long been studied at the electrophysiological level but have lacked the genetic and genomic resources available for research in *Drosophila*. Nonetheless, enough information is available to suggest that certain common principles apply to the coding strategies of different insects. Sphinx moth olfactory neurons are a mixture of narrowly tuned and broadly tuned

cells, with distinct aromatic and aliphatic classes, similar to fly olfactory neurons⁷³. Honeybee antennal lobes have glomeruli that respond broadly to aliphatic compounds, with little specificity for functional groups, similarly to fly glomeruli⁷⁴. Possible differences between insects arise in the patterns of projection and local neuron projections. In moths, some projection neurons collect information from a number of glomeruli^{75,76}, and some local neurons project only to a few glomeruli, whereas others project broadly, like the local neurons of *Drosophila*⁷⁷. A broader view of insect olfaction is provided in the review on page 302.

The detection of natural odours may be sparse

Pure chemicals are rare in nature. Real odours are mixtures of volatiles, and human odour perception is not always dominated by the most abundant organic molecule in a mixture. A deeper understanding of olfaction requires analysis of the complex odours found in the real world. The most extensive study of natural odours so far has been conducted in the mouse, using intrinsic signal imaging of the olfactory bulb. These experiments show that natural concentrations of complex odours, such as that of peanut butter, activate only a small number of glomeruli^{66,78}. Sixty natural food odours, including those of onions, mushrooms, nutmeg and coffee, were mapped; most odours activated less than five glomeruli of the 200 that were investigated⁷⁸. These results suggest that natural odour representations are sparse in the olfactory bulb.

To map the chemical features of complex odours, the individual volatile compounds in coffee and cloves were fractionated by gas chromatography and then tested for activity in mice using intrinsic olfactory imaging. The glomeruli activated by the complex natural odours could be predicted from their responses to about ten individual chemical fractions identified by the gas chromatograph. A mixture of the purified fractions in the correct ratio recapitulated almost all of the olfactory bulb activity observed in response to the complex odour⁷⁸. This analysis suggests that the mouse olfactory system identifies a mixture by responding strongly and specifically to a small number of its constituent chemicals.

From flavour to behaviour

C. elegans has been particularly useful in establishing the link between odour and behaviour. A single *C. elegans* sensory neuron recognizes various chemicals, which is consistent with the expression of multiple receptor genes, but each cell seems to be specialized at the behavioural level⁷⁹. Certain chemosensory neurons recognize attractive tastes or odours; others recognize repulsive tastes or odours; and a third group strongly affect development⁷⁹. These rules are not absolute, but suggest

that individual chemosensory cell types are wired into characteristic behavioural responses. Direct support for this hypothesis has been provided by targeted misexpression of chemosensory receptor genes in inappropriate cell types. The odour diacetyl is attractive to wild-type *C. elegans*, but transgenic misexpression of the diacetyl receptor ODR-10 in AWB neurons, which normally sense repellents, produces animals that avoid diacetyl instead of approaching it⁸⁰. This idea was extended to the expression of a heterologous ion channel, the mammalian transient receptor potential channel TRPV1, in sensory neurons that normally mediate rapid escape behaviours in response to noxious stimuli. Expression of TRPV1 in these neurons caused *C. elegans* to avoid capsaicin, the chilli pepper irritant that causes TRPV1 channels to open⁸¹.

The taste systems of mice and *Drosophila* recapitulate the idea that some chemoreceptor neurons are hard-wired to behavioural responses. In the mouse, bitter receptors and sweet receptors are expressed in non-overlapping epithelial taste cells of the tongue. When a RASSL, a mutated GPCR sensitive to a synthetic opiate, is expressed in the taste cells that normally express bitter receptors, mice reject water flavoured with the opiate⁸². When the same RASSL is expressed in the taste cells that normally express sweet receptors, mice preferentially drink opiate-spiked water⁸³. Similarly, *Drosophila* taste cells for sweet and bitter substances can be activated by capsaicin after transgenic expression of mammalian TRPV1 (ref. 84). If expressed in the sweet-sensing neurons, TRPV1 mediates capsaicin preference; if expressed in the bitter-sensing neurons, TRPV1 mediates capsaicin rejection. These results support the existence of a hard-wired labelled line linking fundamental taste qualities to behavioural outputs.

The results of several studies suggest that at least some olfactory responses are hard-wired and innate. In the fly olfactory system, for example, certain neurons are highly selective for carbon dioxide^{31,32,85}. Stressed flies release carbon dioxide, which acts as an alarm substance to other flies to elicit a strong avoidance response⁸⁵. Carbon-dioxide-sensing neurons are excellent candidates for hard wiring into behavioural responses. The behaviour varies according to species: in mosquitoes, carbon dioxide released by mammalian hosts acts as a strong attractant; the same is true for sphinx moths, which are attracted to carbon-dioxide-producing flowers⁸⁶. Sex-specific olfactory behaviours are also hard-wired into the fly olfactory system^{75,87}.

Mammals, too, have innate responses to certain odours. Predator odours elicit innate defensive freezing or flight responses in mice, even in laboratory mice that have not experienced predation in hundreds of generations⁸⁸. Olfactory neurons that sense predator odours thus seem to be hard-wired into defensive responses. Conversely, odours from potential mates elicit innate sexual responses in mice, largely through the main olfactory system⁸⁹. These olfactory neurons also seem to be hard-wired into behavioural circuits.

Are all olfactory preferences hard-wired? This is unlikely, as even flies and nematodes exhibit olfactory learning^{90,91}. Like the receptors themselves, the behavioural responses to odours are likely to involve many strategies, with hard-wired sexual responses at one end and flexible learned responses at the other extreme.

An evolutionary view of olfaction

A behavioural and ecological view of the olfactory system provides explanations for features that at first seem puzzling. For example, chemosensation is a primitive sense recognizable in unicellular organisms, yet the olfactory system relies on completely independent families of receptor genes in nematodes, insects and vertebrates. This seems a surprising violation of evolutionary conservation, but a rapidly changing olfactory system can be understood as an adaptation to the requirements of a dynamic odour landscape.

About 400 million years ago, nematodes, insects and vertebrates came on land independently and encountered new volatile odours. Departure from the aquatic environment correlates with a massive expansion of odorant receptor genes in vertebrate olfactory systems⁹² — genes almost certainly selected to detect airborne odours. Terrestrial vertebrate odorant receptor genes expanded preferentially from a subset of the fish

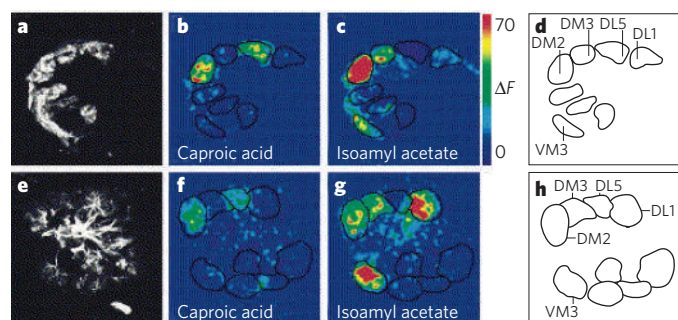


Figure 4 | Odorant sensitivity of olfactory sensory neurons and projection neurons. In antennal lobes, the axons of olfactory sensory neurons and the dendrites of projection neurons can be visualized. **a–d**, Expression of a genetically encoded calcium indicator in a subset of olfactory sensory neurons. **e–h**, Expression of a genetically encoded calcium indicator in a subset of projection neurons. There is considerable overlap in the glomeruli visualized in the upper and lower panels. Odours activate similar sets of sensory neurons and projection neurons. **a** and **e** show the expression patterns of the indicators; **b**, **c**, **f** and **g** show calcium responses after exposure to odours (caproic acid and isoamyl acetate); **d** and **h** provide a glomerular map of the relevant part of the antennal lobe. (Image modified, with permission, from ref. 32.)

odorant receptor repertoire⁹³, and the water-to-air transition can still be detected in the sensory anatomy of amphibians. The 'water nose' that frogs use when swimming preferentially expresses genes similar to those found in fish, whereas the 'air nose' used on land preferentially expresses genes similar to those seen in other terrestrial animals⁹⁴.

At a more specific level, the odours of interest differ among species. Dipteran insects are descended from a common ancestor that lived less than 200 million years ago; each insect species has since evolved its own olfactory preferences and receptor genes. As a result, dipteran fruitflies, mosquitoes, and flesh flies can seek food by tracking volatile metabolites of live plants, live animals or bacterial decomposition, respectively.

Moreover, odours themselves change during evolution. Angiosperms (flowering plants) are the natural source of fruit and flower odours. Angiosperms and insects form the two largest groups of multicellular organisms on Earth. Their diversification about 100 million years ago is thought to have been triggered, on the one hand, by insect pollination, and, on the other, by an arms race between herbivorous insects and the plants they consumed^{95,96}. Thus the many odours produced by fruits and flowers are evolutionarily recent — perhaps 100 million years old. New receptors were needed in insects, vertebrates and nematodes to identify and distinguish these new and informative odours.

The visual system and auditory system are stable because light and sound are immutable physical entities. By contrast, the olfactory system, like the immune system, tracks a moving world of cues generated by other organisms, and must constantly generate, test and discard receptor genes and coding strategies over evolutionary time.

Conclusions and prospects

Much remains to be understood about the nature of odour perception. At the peripheral level, the odour specificity of most mammalian receptors is still unknown. A major opportunity to link biochemistry to perception is provided by human olfaction, in which the manageable number of receptor genes could perhaps be matched to odours. An advantage of working with humans is that they can directly communicate their perception of odour intensity and quality. It is possible that the genetic variation that exists between receptors in different people might someday be matched to individual differences in olfactory perception and preference.

In terms of the molecular biology of mammalian olfaction, each neuron's stochastic choice of a single odorant receptor gene from a large repertoire remains an unsolved mystery. Recent results suggest an unconventional cross-chromosome communication⁹⁷; more will surely be discovered.

The greatest challenge facing the field at present is that of the higher brain levels, where neuronal activity must be assembled into a perception of odour quality. How the combinatorial encoding of odour quality occurs is an important and so far elusive question, and the subject of much ongoing research. Studies of olfaction in *Drosophila* larvae show that the activation of pairs of olfactory neurons triggers behavioural responses that cannot be predicted from the behavioural response observed upon activation of either single neuron⁵⁸. Both flies and mammals reorganize olfactory information considerably at levels above the olfactory bulb or antennal lobe, providing potential anatomical substrates for integration^{98–100}.

Finally, a richer view of olfaction will arise — and perhaps better questions will be asked — as the links between odour, behaviour and ecology are considered and explored. ■

1. Buck, L. & Axel, R. A novel multigene family may encode odorant receptors: a molecular basis for odor recognition. *Cell* **65**, 175–187 (1991).
2. Niimura, Y. & Nei, M. Comparative evolutionary analysis of olfactory receptor gene clusters between humans and mice. *Gene* **14**, 13–21 (2005).
3. Niimura, Y. & Nei, M. Evolutionary changes of the number of olfactory receptor genes in the human and mouse lineages. *Gene* **346**, 23–28 (2005).
4. Robertson, H. A. & Thomas, J. H. The putative chemoreceptor families of *C. elegans*. *Wormbook* <<http://www.wormbook.org>> doi/10.1895/wormbook.1.66.1 (2006).
5. Troemel, E. R., Chou, J. H., Dwyer, N. D., Colbert, H. A. & Bargmann, C. I. Divergent seven transmembrane receptors are candidate chemosensory receptors in *C. elegans*. *Cell* **83**, 207–218 (1995).

6. Colosimo, M. E. *et al.* Identification of thermosensory and olfactory neuron-specific genes via expression profiling of single neuron types. *Curr. Biol.* **14**, 2245–2251 (2004).
7. McCarroll, S. A., Li, H. & Bargmann, C. I. Identification of transcriptional regulatory elements in chemosensory receptor genes by probabilistic segmentation. *Curr. Biol.* **15**, 347–352 (2005).
8. Chen, N. *et al.* Identification of a nematode chemosensory gene family. *Proc. Natl Acad. Sci. USA* **102**, 146–151 (2005).
9. Gray, J. M. *et al.* Oxygen sensation and social feeding mediated by a *C. elegans* guanylate cyclase homologue. *Nature* **430**, 317–322 (2004).
10. Cheung, B. H., Cohen, M., Rogers, C., Albayram, O. & de Bono, M. Experience-dependent modulation of *C. elegans* behavior by ambient oxygen. *Curr. Biol.* **15**, 905–917 (2005).
11. Yu, S., Avery, L., Baude, E. & Garbers, D. L. Guanylyl cyclase expression in specific sensory neurons: a new family of chemosensory receptors. *Proc. Natl Acad. Sci. USA* **94**, 3384–3387 (1997).
12. Clyne, P. J. *et al.* A novel family of divergent seven-transmembrane proteins: candidate odorant receptors in *Drosophila*. *Neuron* **22**, 327–338 (1999).
13. Vosshall, L. B., Amrein, H., Morozov, P. S., Rzhetsky, A. & Axel, R. A spatial map of olfactory receptor expression in the *Drosophila* antenna. *Cell* **96**, 725–736 (1999).
14. Benton, R., Sachse, S., Michnick, S. W. & Vosshall, L. B. Atypical membrane topology and heteromeric function of *Drosophila* odorant receptors *in vivo*. *PLoS Biol.* **4**, e20 (2006).
15. Wistrand, M., Kall, L. & Sonnhämmer, E. L. A general model of G protein-coupled receptor sequences and its application to detect remote homologs. *Protein Sci.* **15**, 509–521 (2006).
16. Neuhaus, E. M. *et al.* Odorant receptor heterodimerization in the olfactory system of *Drosophila melanogaster*. *Nature Neurosci.* **8**, 15–17 (2005).
17. Zhao, H. *et al.* Functional expression of a mammalian odorant receptor. *Science* **279**, 237–242 (1998).
18. Krautwurst, D., Yau, K. W. & Reed, R. R. Identification of ligands for olfactory receptors by functional expression of a receptor library. *Cell* **95**, 917–926 (1998).
19. Malnic, B., Hirono, J., Sato, T. & Buck, L. B. Combinatorial receptor codes for odors. *Cell* **96**, 713–723 (1999).
20. Wetzel, C. H. *et al.* Specificity and sensitivity of a human olfactory receptor functionally expressed in human embryonic kidney 293 cells and *Xenopus laevis* oocytes. *J. Neurosci.* **19**, 7426–7433 (1999).
21. Kajiyama, K. *et al.* Molecular bases of odor discrimination: reconstitution of olfactory receptors that recognize overlapping sets of odorants. *J. Neurosci.* **21**, 6018–6025 (2001).
22. Touhara, K. *et al.* Functional identification and reconstitution of an odorant receptor in single olfactory neurons. *Proc. Natl Acad. Sci. USA* **96**, 4040–4045 (1999).
23. Araneda, R. C., Kini, A. D. & Firestein, S. The molecular receptive range of an odorant receptor. *Nature Neurosci.* **3**, 1248–1255 (2000).
24. Spehr, M. *et al.* Identification of a testicular odorant receptor mediating human sperm chemotaxis. *Science* **299**, 2054–2058 (2003).
25. Gether, U. & Kobilka, B. K. G protein-coupled receptors. II. Mechanism of agonist activation. *J. Biol. Chem.* **273**, 17979–17982 (1998).
26. Palczewski, K. *et al.* Crystal structure of rhodopsin: a G protein-coupled receptor. *Science* **289**, 739–745 (2000).
27. Lu, Z. L., Saldanha, J. W. & Hulme, E. C. Seven-transmembrane receptors: crystals clarify. *Trends Pharmacol. Sci.* **23**, 140–146 (2002).
28. Katada, S., Hirokawa, T., Oka, Y., Suwa, M. & Touhara, K. Structural basis for a broad but selective ligand spectrum of a mouse olfactory receptor: mapping the odorant-binding site. *J. Neurosci.* **25**, 1806–1815 (2005).
29. Oka, Y., Omura, M., Kataoka, H. & Touhara, K. Olfactory receptor antagonism between odorants. *EMBO J.* **14**, 120–126 (2004).
30. de Bruyne, M., Clyne, P. J. & Carlson, J. R. Odor coding in a model olfactory organ: the *Drosophila* maxillary palp. *J. Neurosci.* **19**, 4520–4532 (1999).
31. de Bruyne, M., Foster, K. & Carlson, J. R. Odor coding in the *Drosophila* antenna. *Neuron* **30**, 537–552 (2001).
32. Wang, J. W., Wong, A. M., Flores, J., Vosshall, L. B. & Axel, R. Two-photon calcium imaging reveals an odor-evoked map of activity in the fly brain. *Cell* **112**, 271–282 (2003).
33. Elmore, T., Ignell, R., Carlson, J. R. & Smith, D. P. Targeted mutation of a *Drosophila* odor receptor defines receptor requirement in a novel class of sensillum. *J. Neurosci.* **23**, 9906–9912 (2003).
34. Yao, C. A., Ignell, R. & Carlson, J. R. Chemosensory coding by neurons in the coeloconic sensilla of the *Drosophila* antenna. *J. Neurosci.* **25**, 8539–8567 (2005).
35. Hallem, E. A., Ho, M. G. & Carlson, J. R. The molecular basis of odor coding in the *Drosophila* antenna. *Cell* **117**, 965–979 (2004).
36. Kreher, S. A., Kwon, J. Y. & Carlson, J. R. The molecular basis of odor coding in the *Drosophila* larva. *Neuron* **46**, 445–456 (2005).
37. Hallem, E. A. & Carlson, J. R. Coding of odors by a receptor repertoire. *Cell* **125**, 143–160 (2006).
38. Sengupta, P., Chou, J. C. & Bargmann, C. I. *odr-10* encodes a seven transmembrane domain olfactory receptor required for responses to the odorant diacetyl. *Cell* **84**, 899–909 (1996).
39. Chess, A., Simon, I., Cedar, H. & Axel, R. Allelic inactivation regulates olfactory receptor gene expression. *Cell* **78**, 823–834 (1994).
40. Serizawa, S. *et al.* Negative feedback regulation ensures the one receptor—one olfactory neuron rule in mouse. *Science* **302**, 2088–2094 (2003).
41. Ressler, K. J., Sullivan, S. L. & Buck, L. B. A zonal organization of odorant receptor gene expression in the olfactory epithelium. *Cell* **73**, 597–609 (1993).
42. Vassar, R., Ngai, J. & Axel, R. Spatial segregation of odorant receptor expression in the mammalian olfactory epithelium. *Cell* **74**, 309–318 (1993).
43. Qasba, P. & Reed, R. R. Tissue and zonal-specific expression of an olfactory receptor transgene. *J. Neurosci.* **18**, 227–236 (1998).
44. Lewcock, J. W. & Reed, R. R. A feedback mechanism regulates monoallelic odorant receptor expression. *Proc. Natl Acad. Sci. USA* **101**, 1069–1074 (2004).
45. Shykind, B. M. *et al.* Gene switching and the stability of odorant receptor gene choice. *Cell* **117**, 801–815 (2004).
46. Larsson, M. C. *et al.* *Or83b* encodes a broadly expressed odorant receptor essential for *Drosophila* olfaction. *Neuron* **43**, 703–714 (2004).

47. Fishilevich, E. & Vosshall, L. B. Genetic and functional subdivision of the *Drosophila* antennal lobe. *Curr. Biol.* **15**, 1548–1553 (2005).
48. Couto, A., Alenius, M. & Dickson, B. J. Molecular, anatomical, and functional organization of the *Drosophila* olfactory system. *Curr. Biol.* **15**, 1535–1547 (2005).
49. Dobritsa, A. A. *et al.* Integrating the molecular and cellular basis of odor coding in the *Drosophila* antenna. *Neuron* **37**, 827–841 (2003).
50. Goldman, A. L., Van der Goes van Naters, W., Lessing, D., Warr, C. G. & Carlson, J. R. Coexpression of two functional odor receptors in one neuron. *Neuron* **45**, 661–666 (2005).
51. Sengupta, P., Colbert, H. A. & Bargmann, C. I. The *C. elegans* gene *odr-7* encodes an olfactory-specific member of the nuclear receptor superfamily. *Cell* **79**, 971–980 (1994).
52. Lanjuin, A. & Sengupta, P. Specification of chemosensory neuron subtype identities in *Caenorhabditis elegans*. *Curr. Opin. Neurobiol.* **14**, 22–30 (2004).
53. Johnston, R. J. J., Chang, S., Etchberger, J. F., Ortiz, C. O. & Hobert, O. MicroRNAs acting in a double-negative feedback loop to control a neuronal cell fate decision. *Proc. Natl Acad. Sci. USA* **102**, 12449–12454 (2005).
54. Troemel, E. R., Sagasti, A. & Bargmann, C. I. Lateral signaling mediated by axon contact and calcium entry regulates asymmetric odorant receptor expression in *C. elegans*. *Cell* **99**, 387–398 (1999).
55. Peckol, E. L., Troemel, E. R. & Bargmann, C. I. Sensory experience and sensory activity regulate chemosensory receptor gene expression in *Caenorhabditis elegans*. *Proc. Natl Acad. Sci. USA* **98**, 11032–11038 (2001).
56. Nolan, K. M., Sarafi-Reinach, T. R., Horne, J. G., Saffer, A. M. & Sengupta, P. The DAF-7 TGF- β signaling pathway regulates chemosensory receptor gene expression in *C. elegans*. *Genes Dev.* **16**, 3061–3073 (2002).
57. Fox, A. N., Pitts, R. J., Robertson, H. M., Carlson, J. R. & Zwiebel, L. J. Candidate odorant receptors from the malaria vector mosquito *Anopheles gambiae* and evidence of down-regulation in response to blood feeding. *Proc. Natl Acad. Sci. USA* **98**, 14693–14697 (2001).
58. Fishilevich, E. *et al.* Chemotaxis behavior mediated by single larval olfactory neurons in *Drosophila*. *Curr. Biol.* **15**, 2086–2096 (2005).
59. Barth, A. L., Justice, N. J. & Ngai, J. Asynchronous onset of odorant receptor expression in the developing zebrafish olfactory system. *Neuron* **16**, 23–34 (1996).
60. Ressler, K. J., Sullivan, S. L. & Buck, L. B. Information coding in the olfactory system: evidence for a stereotyped and highly organized epitope map in the olfactory bulb. *Cell* **79**, 1245–1256 (1994).
61. Vassar, R. *et al.* Topographic organization of sensory projections to the olfactory bulb. *Cell* **79**, 981–992 (1994).
62. Strotmann, J. *et al.* Local permutations in the glomerular array of the mouse olfactory bulb. *J. Neurosci.* **20**, 6927–6938 (2000).
63. Komiya, T. & Luo, L. Development of wiring specificity in the olfactory system. *Curr. Opin. Neurobiol.* **16**, 67–73 (2006).
64. White, J. G., Southgate, E., Thomson, J. N. & Brenner, S. The structure of the nervous system of the nematode *Caenorhabditis elegans*. *Phil. Trans. R. Soc. Lond. B* **314**, 1–340 (1986).
65. Mori, K., Takahashi, Y. K., Igarashi, K. M. & Yamaguchi, M. Maps of odorant molecular features in the mammalian olfactory bulb. *Physiol. Rev.* **86**, 409–433 (2006).
66. Rubin, B. & Katz, L. Optical imaging of odorant representations in the mammalian olfactory bulb. *Neuron* **23**, 499–511 (1999).
67. Meister, M. & Bonhoeffer, T. Tuning and topography in an animal olfactory bulb. *J. Neurosci.* **15**, 1351–1360 (2001).
68. Baier, H. & Korsching, S. Olfactory glomeruli in the zebrafish form an invariant pattern and are identifiable across animals. *J. Neurosci.* **14**, 219–230 (1994).
69. Friedrich, R. W. & Korsching, S. I. Combinatorial and chemotopic odorant coding in the zebrafish olfactory bulb visualized by optical imaging. *Neuron* **18**, 737–752 (1997).
70. Wilson, R. I., Turner, G. C. & Laurent, G. Transformation of olfactory representations in the *Drosophila* antennal lobe. *Science* **303**, 366–370 (2004).
71. Wilson, R. I. & Laurent, G. Role of GABAergic inhibition in shaping odor-evoked spatiotemporal patterns in the *Drosophila* antennal lobe. *J. Neurosci.* **25**, 9069–9079 (2005).
72. Ng, M. *et al.* Transmission of olfactory information between three populations of neurons in the antennal lobe of the fly. *Neuron* **36**, 463–474 (2002).
73. Shields, V. D. & Hildebrand, J. G. Responses of a population of antennal olfactory receptor cells in the female moth *Manduca sexta* to plant-associated volatile organic compounds. *J. Comp. Physiol. A* **186**, 1135–1151 (2001).
74. Galizia, C. G. & Menzel, R. The role of glomeruli in the neural representation of odours: results from optical recording studies. *J. Insect Physiol.* **47**, 115–130 (2001).
75. Christensen, T. A. & Hildebrand, J. G. Male-specific, sex pheromone-selective projection neurons in the antennal lobes of the moth *Manduca sexta*. *J. Comp. Physiol. A* **160**, 553–569 (1987).
76. Hansson, B. S., Christensen, T. A. & Hildebrand, J. G. Functionally distinct subdivisions of the macroglomerular complex in the antennal lobe of the male sphinx moth *Manduca sexta*. *J. Comp. Neurol.* **312**, 264–278 (1991).
77. Christensen, T. A., Waldrop, B. R., Harrow, I. D. & Hildebrand, J. G. Local interneurons and information processing in the olfactory glomeruli of the moth *Manduca sexta*. *J. Comp. Physiol. A* **173**, 385–399 (1993).
78. Lin, D. Y., Shea, S. D. & Katz, L. C. Representation of natural stimuli in the rodent main olfactory bulb. *Neuron* **50**, 937–949 (2006).
79. Bargmann, C. I. & Mori, I. in *C. elegans II* (eds Riddle, D. L., Blumenthal, T., Meyer, B. J. & Priess, J. R.) 717–737 (Cold Spring Harbor Laboratory Press, New York, 1997).
80. Troemel, E. R., Kimmel, B. E. & Bargmann, C. I. Reprogramming chemotaxis responses: sensory neurons define olfactory preferences in *C. elegans*. *Cell* **91**, 161–169 (1997).
81. Tobin, D. *et al.* Combinatorial expression of TRPV channel proteins defines their sensory functions and subcellular localization in *C. elegans* neurons. *Neuron* **35**, 307–318 (2002).
82. Mueller, K. L. *et al.* The receptors and coding logic for bitter taste. *Nature* **434**, 225–229 (2005).
83. Zhao, G. Q. *et al.* The receptors for mammalian sweet and umami taste. *Cell* **115**, 255–266 (2003).
84. Marella, S. *et al.* Imaging taste responses in the fly brain reveals a functional map of taste category and behavior. *Neuron* **49**, 285–295 (2006).
85. Suh, G. S. *et al.* A single population of olfactory sensory neurons mediates an innate avoidance behaviour in *Drosophila*. *Nature* **431**, 854–859 (2004).
86. Thom, C., Guerenstein, P. G., Mechaber, W. L. & Hildebrand, J. G. Floral CO₂ reveals flower profitability to moths. *J. Chem. Ecol.* **30**, 1285–1288 (2004).
87. Stockinger, P., Kvitsiani, D., Rotkopf, S., Tirian, L. & Dickson, B. J. Neural circuitry that governs *Drosophila* male courtship behavior. *Cell* **121**, 795–807 (2005).
88. Mongeau, R., Miller, G. A., Chiang, E. & Anderson, D. J. Neural correlates of competing fear behaviors evoked by an innately aversive stimulus. *J. Neurosci.* **23**, 3855–3868 (2003).
89. Mandiyan, V. S., Coats, J. K. & Shah, N. M. Deficits in sexual and aggressive behaviors in *Cnga2* mutant mice. *Nature Neurosci.* **8**, 1660–1662 (2005).
90. Quinn, W. G., Harris, W. A. & Benzer, S. Conditioned behavior in *Drosophila melanogaster*. *Proc. Natl Acad. Sci. USA* **71**, 708–712 (1974).
91. Zhang, Y., Lu, H. & Bargmann, C. Pathogenic bacteria induce aversive olfactory learning in *Caenorhabditis elegans*. *Nature* **438**, 179–184 (2005).
92. Glusman, G. *et al.* The olfactory receptor gene superfamily: data mining, classification, and nomenclature. *Mamm. Genome* **11**, 1016–1023 (2000).
93. Niimura, Y. & Nei, M. Evolutionary dynamics of olfactory receptor genes in fishes and tetrapods. *Proc. Natl Acad. Sci. USA* **102**, 6039–6044 (2005).
94. Freitag, J., Krieger, J., Strotmann, J. & Breer, H. Two classes of olfactory receptors in *Xenopus laevis*. *Neuron* **15**, 1383–1392 (1995).
95. Crane, P. R., Friis, E. M. & Pedersen, K. R. The origin and early diversification of angiosperms. *Nature* **374**, 27–33 (1995).
96. Bernays, E. A. Evolution of feeding behavior in insect herbivores. *Bioscience* **48**, 35–44 (1998).
97. Lomvardas, S. *et al.* Interchromosomal interactions and olfactory receptor choice. *Cell* **126**, 403–413 (2006).
98. Marin, E. C., Jefferis, G. S., Komiya, T., Zhu, H. & Luo, L. Representation of the glomerular olfactory map in the *Drosophila* brain. *Cell* **109**, 243–255 (2002).
99. Wong, A. M., Wang, J. W. & Axel, R. Spatial representation of the glomerular map in the *Drosophila* protocerebrum. *Cell* **109**, 229–241 (2002).
100. Zou, Z., Horowitz, L. F., Montmayeur, J. P., Snapper, S. & Buck, L. B. Genetic tracing reveals a stereotyped sensory map in the olfactory cortex. *Nature* **414**, 173–179 (2001).

Acknowledgements R. Wilson and L. Vosshall greatly improved this piece by their comments and corrections. C.I.B. is an Investigator of the Howard Hughes Medical Institute; the laboratory's work on olfaction is funded by the NIDCD.

Author information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The author declares no competing financial interests. Correspondence should be addressed to the author (cori@rockefeller.edu).

Insects as chemosensors of humans and crops

Wynand van der Goes van Naters¹ & John R. Carlson¹

Insects transmit disease to hundreds of millions of people a year, and cause enormous losses to the world's agricultural output. Many insects find the human or plant hosts on which they feed, and identify and locate their mates, primarily through olfaction and taste. Major advances have recently been made in understanding insect chemosensation at the molecular and cellular levels. These advances have provided new opportunities to control insects that cause massive damage to health and agriculture across the world.

The problem of insect pests is not a new one, and there are many historical examples of the devastation they can inflict on both human health and agriculture worldwide. Over time, various efforts have been made to control such pests, but their efficacy has been limited and none has been without drawbacks. Recent progress in the molecular and cellular biology of insect olfaction and taste offers opportunities for the development of new means of pest control. We believe that new strategies based on a molecular understanding of insect chemosensation may, in conjunction with existing strategies, contribute to the control of insects that cause enormous harm to human health and agriculture.

Insect olfaction and disease

Insects carry a wide variety of diseases, including malaria (Fig. 1), dengue fever, yellow fever, trypanosomiasis (African sleeping sickness and Chagas disease), plague, lymphatic filariasis (elephantiasis), onchocerciasis (river blindness), leishmaniasis, and some forms of encephalitis and typhus. In these diseases, the insect acts as a vector for a parasite — the mosquito *Anopheles gambiae* is a major vector of the plasmodia that cause malaria, and tsetse flies transmit trypanosomes that cause sleeping sickness. Malaria alone is responsible for more than 500 million clinical disease episodes per year¹, and 3.2 billion people — half

the world's population — are at risk². Each year, malaria causes more than a million deaths, mostly in children in Africa. Although many insect-borne diseases are primarily tropical and subtropical, the recent emergence of West Nile encephalitis in the United States illustrates their occurrence in temperate climates.

Human diseases are not the only threat to the world's population — insects also transmit many diseases to animals, including livestock, with serious economic consequences. Tsetse flies, for example, spread a form of trypanosomiasis called nagana that has had devastating effects on the cattle of sub-Saharan Africa.

Many insect vectors of disease find their human hosts primarily through olfactory cues. Female *A. gambiae* are sensitive to various compounds emanated by humans — CO₂, lactic acid, ammonia, carboxylic acids and other compounds — some of which act synergistically to promote an attractive response³. Individual humans vary in their attractiveness to mosquitoes, and there is evidence that these differences depend at least in part on variations in body odour^{4,5}. Although *A. gambiae* is anthropophilic⁶ — that is, it feeds primarily on human hosts — its sibling species often feed on both human and animal hosts, and the anthropophily of *A. gambiae* is believed to have an olfactory basis⁷.

Insect chemosensation and agriculture

Approximately 35% of all agricultural crops are lost to pests — primarily insects and plant pathogens, which insects often transmit⁸. Major world food crops such as wheat, rice, maize and potatoes are all susceptible to damage by insects such as the Hessian fly (wheat), the green leafhopper (rice), the corn-borer moth and the Colorado potato beetle (Fig. 2). Fruit and vegetable crops also suffer attack from a wide variety of taxonomically diverse insects, as do other crops such as cotton. Between 1892 and 1999, the cotton boll weevil cost the US cotton industry \$17 billion (unadjusted for inflation) as a result of both crop losses due to damage and the cost of insecticide and its application⁹. Stored agricultural products such as grains, cereal products, nuts and dried beans undergo heavy losses due to insects such as weevils, and great expenditure is necessary to prevent additional losses. Much damage is caused directly by adult insects, but in the case of many agricultural pests the greatest losses follow the deposition of eggs on crops. After hatching, the larvae or juveniles consume or injure the host plant. Insects such as aphids also contribute to crop losses by transmitting viruses and other pathogens to plants¹⁰.

Insects locate and select their host plants largely on the basis of chemosensory cues (Box 1). Olfactory cues can mediate long-range attraction.



Figure 1 | An insect vector of human disease. The mosquito *Anopheles minimus* is a vector of malaria. (Image courtesy of James Gathany/CDC.)

¹Department of Molecular, Cellular and Developmental Biology, Yale University, PO Box 208103, New Haven, Connecticut 06520-8103, USA.



Figure 2 | Crop plants are vulnerable to insect pests. Colorado potato beetles feeding on the leaf of a potato plant. (Image courtesy of J. Dickens and B. Hollister, Agricultural Research Service, United States Department of Agriculture, Beltsville, Maryland.)

For example, the onion fly *Delia antiqua* is attracted to the host volatile dipropyl disulphide from distances of 100 m or more¹¹. *Rhagoletis* flies, which are pests of apple orchards, show an attraction response to apple odours, and a five-component blend of apple volatiles elicits attractive responses in laboratory and field tests¹². Plant taste cues can either stimulate or inhibit feeding, and can also affect egg laying by females. *Manduca sexta* caterpillars, for example, are stimulated to feed on their host plants by some compounds, including various sugars, and are deterred by others, including caffeine and aristolochic acid¹³, which taste bitter to humans. Some insects have highly specialized chemosensory neurons that respond to a taxon-specific plant compound that stimulates feeding or oviposition in these insects, and is, in some cases,

necessary for acceptance^{14,15}. For example, cucurbitacins (a group of oxygenated tetracyclic triterpenes) act as feeding stimulants for some diabrotic beetle species that include serious pests such as the western corn rootworm *Diabrotica virgifera*¹⁶.

Chemical communication within insect pest species

Many insect pests communicate with others of the same species through pheromones, molecules that are secreted by one individual and that elicit a response in another of the same species. Some pheromones function as attractants, allowing individuals to detect and locate mates, whereas others act as alarms, induce trail-following, or influence oviposition or aggregation behaviour (the tendency of individuals to congregate).

Pheromones have been identified from hundreds of insect species¹⁷, and many are long-chain unsaturated acetate esters, alcohols or aldehydes. Female moths typically emit a blend of one or two major components and several minor components. Species specificity is conferred both by the identity of the components and by their proportions in the mixture¹⁸.

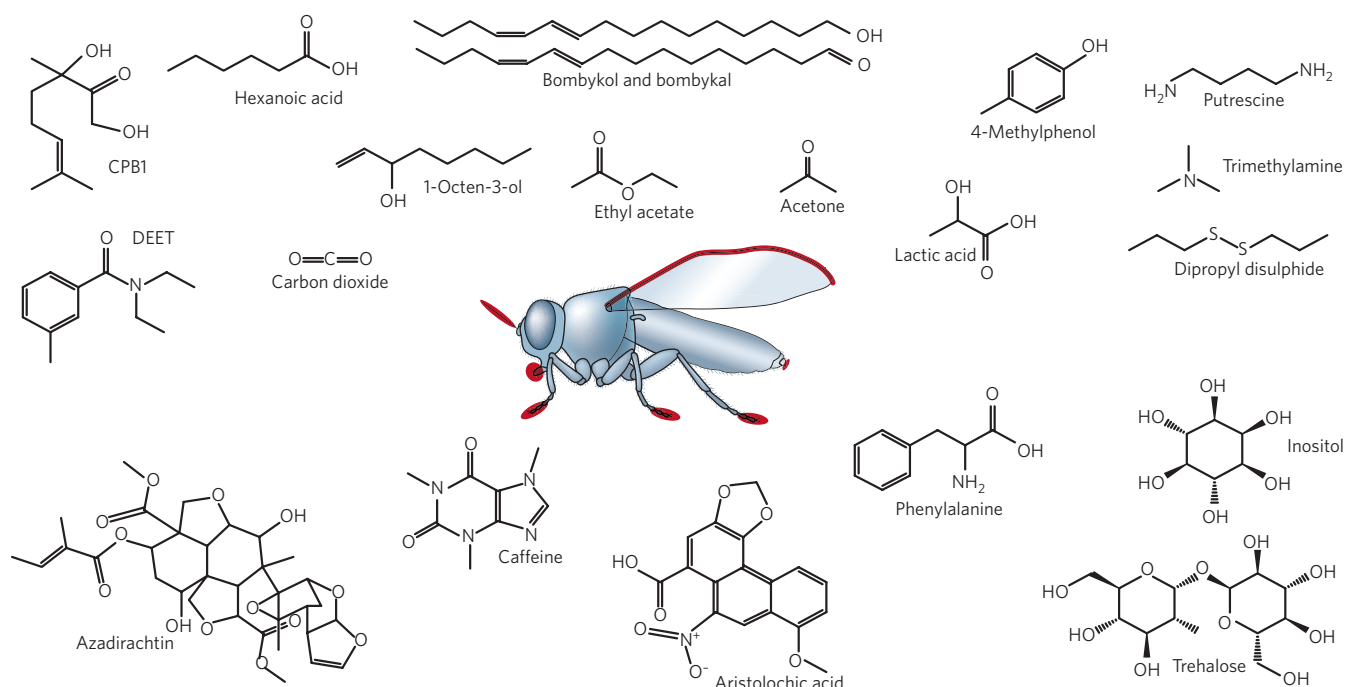
The first sex attractant to be identified in animals was one secreted by female moths, which are able to attract males from a distance. The aliphatic C₁₆ alcohol, known as bombykol ((*E,Z*)-10,12-hexadecadien-1-ol), was isolated from the silk moth *Bombyx mori*. Females release bombykol from the tip of their abdomens, and the chemical elicits wing fluttering and upwind movement in the male¹⁹. The identification of this pheromone spurred incisive studies into the cellular basis of olfaction.

The cellular basis of insect chemosensation

Insects are literally bristling with chemosensory organs. Sensory neurons for olfaction and taste are housed inside bristles known as sensilla. Chemosensory sensilla are found on antennae, mouthparts, legs, wings and genital regions, including ovipositors, from which eggs are laid. Most of an insect's olfactory receptor neurons (ORNs) are located in the antennae, whereas its gustatory neurons are highly concentrated in the mouthparts and legs. Odorants or tastants enter the sensilla through pores and then pass through a lymph to reach the dendrites — extensions of the neurons in which signal transduction is thought to occur.

Box 1 | The chemical world of an insect

A selection of odorants are indicated above the insect, tastants below. Red regions indicate principal locations of chemosensory sensilla.



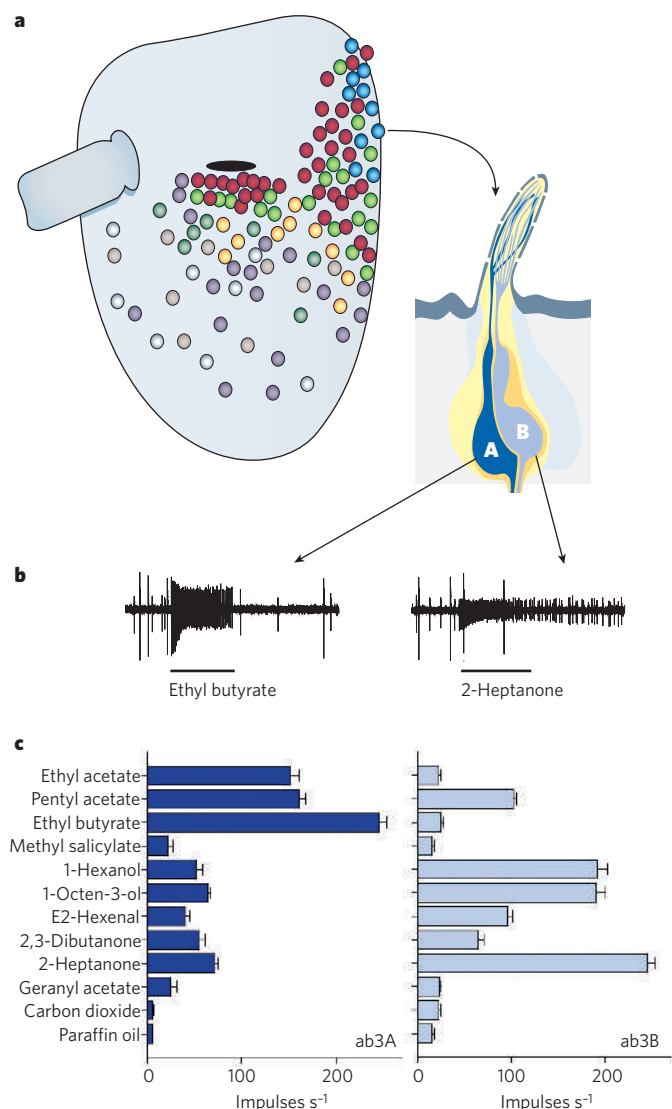


Figure 3 | Anatomy and physiology of the *Drosophila melanogaster* antenna. **a**, The left panel shows the posterior face of the antenna, with coloured dots representing sensilla of different functional types (images adapted, with permission, from refs 52 and 79). Each sensillum contains multiple ORNs, identified as A and B in the magnified sensillum in the right panel. **b**, The electrophysiological recordings show excitatory responses of two individual ORNs, ab3A (large action potentials) and ab3B (small action potentials) neurons to vapours from 10^{-5} dilutions of the indicated odorants. Bars indicate 0.5 s odor pulses. Note that the response to ethyl butyrate is of shorter duration than that to 2-heptanone. **c**, Response profiles to a panel of odorants (10^{-2} dilutions). (Data from ref. 49.)

The activities of antennal ORNs have been measured in electrophysiological recordings from moths, cockroaches, weevils and other insects. These studies have revealed that ORNs differ widely in their odour-response spectra as well as in the temporal dynamics and mode (excitation versus inhibition) of the responses they show to particular odorants^{20,21} (Fig. 3). Some ORNs respond broadly to a variety of odorants, whereas others seem to be narrowly tuned — for example, to a particular pheromone component.

Biochemical studies of the moth *Antheraea polyphemus* identified an abundant pheromone-binding protein (PBP) in the sensillum lymph of the male antenna²². The structure of a PBP from the moth *B. mori* has been defined^{23,24}. On the basis of biochemical and physiological data^{25,26}, a model of its mechanism of action has been proposed^{27,28}: PBP binds a pheromone and then transports it to the ORN dendrites, where the PBP

undergoes a conformational change and the pheromone is released to a receptor. Receptor activation leads to the generation of action potentials, which are transmitted to the brain by the ORN axon.

Physiological recordings from taste sensilla have likewise revealed many types of taste neuron. In some caterpillars, sugars, amino acids and inositol activate separate taste neurons²⁹. Caterpillars also have 'deterrent' cells that respond to an array of alkaloid and phenolic plant compounds that inhibit feeding³⁰. Tsetse flies have taste neurons on their mouthparts that respond to ATP³¹, a stimulus that is sufficient to cause the fly to imbibe a solution isotonic with blood. The leg sensilla of tsetse flies contain taste neurons that respond to uric acid and certain amino acids, which are also positive stimuli for feeding^{32,33}.

Classic strategies of chemosensory-based insect control

Diverse strategies have been used in attempts to control insect pests. In addition to the use of insecticides to kill them, the chemosensory systems of insect pests have been targeted with the goal of repelling them, attracting them to traps or confusing them.

Insect repellents have long been used to protect humans against insect pests. The most widely used repellent at present is DEET (*N,N*-diethyl-3-methylbenzamide)³⁴. Despite its widespread use, its mechanism of action is unknown. Many repellents of natural origin are also used, but in most cases their effectiveness has not yet been rigorously evaluated.

Trapping has been used successfully to control insect pests. Olfactory attractants are used to lure insects to a trap, which often contains an insecticide. Tsetse flies have been the focus of extensive trapping programmes in Africa. Analysis of volatiles from cattle and their urine identified CO₂, acetone, 1-octen-3-ol, 3-n-propylphenol and 4-methylphenol as components that elicit behavioural activity, and blends of some of these compounds have been used in conjunction with visual cues to attract tsetse flies to insecticide-laden screens³⁵.

Attraction to products of protein degradation forms the basis of trap lures like those that target female pest flies such as the Mediterranean fruit fly (or medfly) *Ceratitis capitata*, which is one of the most serious fruit and vegetable insect pests worldwide. One commercially available lure consists of ammonium acetate, putrescine and trimethylamine, and has been effective both for monitoring fly populations and in mass trapping.

Pheromones have also been used effectively as trapping agents. The Colorado potato beetle, *Leptinotarsa decemlineata*, is one of the most important pests affecting potato crops in the United States and Europe. Males of this species produce an aggregation pheromone, (S)-3,7-dimethyl-2-oxo-oct-6-ene-1,3-diol (CPB1), which is being investigated as an attractant for control of this pest³⁶. In addition, traps baited with pheromones are crucial in integrated pest management as a means of monitoring population levels of pest species. The timing and intensity of control measures are often guided by trapping data.

Pheromones have also been used to disrupt mating. When released in large amounts from many dispensers they can act as 'confusants', presumably preventing males from locating calling females. In some cases, the release of a single component of a pheromone blend is effective, possibly because it distorts the ratio of components. This strategy has been used successfully — for example, in the control of the European grape-berry moth *Eupoecilia ambiguella* in vineyards³⁷.

Antifeedants are aversive compounds that prevent insects from feeding on plants. A classic example is azadirachtin, a compound found in the neem tree³⁸. Plagues of desert locusts (*Schistocerca gregaria*) have been observed to defoliate almost all of the green plants they encounter with the exception of the neem tree, and azadirachtin has been identified as a compound that inhibits feeding. Subsequent studies showed that when azadirachtin is applied to other plants it inhibits feeding by locusts on them as well. So far, almost 900 compounds have been found that deter feeding in various insect species³⁹.

Although these four chemosensory-based strategies have been useful in controlling insect pests, they all have limitations. For example, DEET is of limited efficacy, most notably against *Anopheles albimanus*, the principal vector of malaria in Central America and the Caribbean⁴⁰.

Many compounds, including pheromones, are expensive to produce and are labile under field conditions. Because of such limitations, the currently available means of chemosensory-based pest control are used less commonly than other means.

The use of insecticides is the most prevalent method of pest control. However, an important concern is that many pest species have developed insecticide resistance. In many locations, for example, the Colorado potato beetle is now resistant to almost all available synthetic insecticides⁴¹. Another concern is the environmental burden of widespread insecticide use, including the effects on non-target species. Finally, the rising organic agriculture movement demands insecticide-residue-free food.

An alternative strategy is the use of Bt-crops, which are engineered to carry one or more toxin genes from the insect bacterial pathogen *Bacillus thuringiensis*. These toxins are believed to bind receptors in the insect gut and create leakage channels in cell membranes. Susceptible insects die of starvation and septicaemia⁴². Recently, Bt-crops have begun to replace traditional cultivars at a spectacular rate. However, the durability of the conferred protection, the effect of introgression into wild relatives of crop plants, and the long-term effect of Bt-crop plants on non-target insects and the ecosystem as a whole are as yet not clear.

Another means of pest control, the sterile insect technique (SIT), has been very successful against a number of pest species including the screw-worm fly *Cochliomyia hominivorax*, whose larvae feed in the flesh of livestock. SIT involves rearing large numbers of an insect species, and, after irradiating them to cause sterility, releasing them into the pest population; sterilized males can compete with native males for females but do not produce offspring, thereby reducing the population size in the next generation.

An additional strategy is that of biological control — the introduction or augmentation of predators, parasitoids and insect pathogens directed against the pest species. Of 5,576 attempted introductions of natural enemy arthropods to insect pest species, 625 (~11%) have yielded substantial or complete control, although there is a risk of unintended environmental consequences, including the risk that the introduced species might attack non-target hosts⁴³. Clearly there is a need for other strategies that are effective, economical and environmentally benign.

Molecular genetics of insect olfaction and taste

Most of the chemosensory-based control strategies described above were designed without the benefit of a molecular understanding of insect chemosensation. DEET³⁴, for example, was identified in a large-scale screen of compounds and was patented by the US Army in 1946. Recently, there have been major advances in our understanding of insect olfaction and taste, advances that could be exploited to design new strategies.

Insect odorant receptor genes were identified in *Drosophila* in 1999 using computational and molecular approaches^{44,45}. *D. melanogaster* has a family of 60 odorant receptor (*Or*) genes⁴⁶ (Fig. 4), with different genes expressed in different subsets of ORNs. Their identity as odorant receptor genes was confirmed by the finding that one odorant receptor gene conferred a response to several odors when overexpressed in the antenna⁴⁷ and when expressed in *Xenopus* oocytes⁴⁸. Subsequently, a mutant lacking the *Or22a* gene was found to lack odour response in one particular ORN⁴⁹. This mutant ORN, known as 'the empty neuron', was then used as a laboratory in which to express other odorant receptors^{49,50}. The response properties conferred by particular receptors were found to agree closely with the response properties of particular ORNs, thereby validating the fidelity of the expression system. The empty neuron was then used to characterize in detail the response properties of 37 *Drosophila* odorant receptors — most of the other adult odorant receptors as well as many of those of the larva^{50–53}.

The sequences of the *Drosophila* odorant receptor genes have been used to identify orthologues in various insect species. The first orthologues were identified in *A. gambiae*⁵⁴, which has 79 AgOr (*Anopheles gambiae* odour receptor) genes⁵⁵. *Or* orthologues have also been identified in the moth *Heliothis virescens*⁵⁶, a major pest of crops such as soya

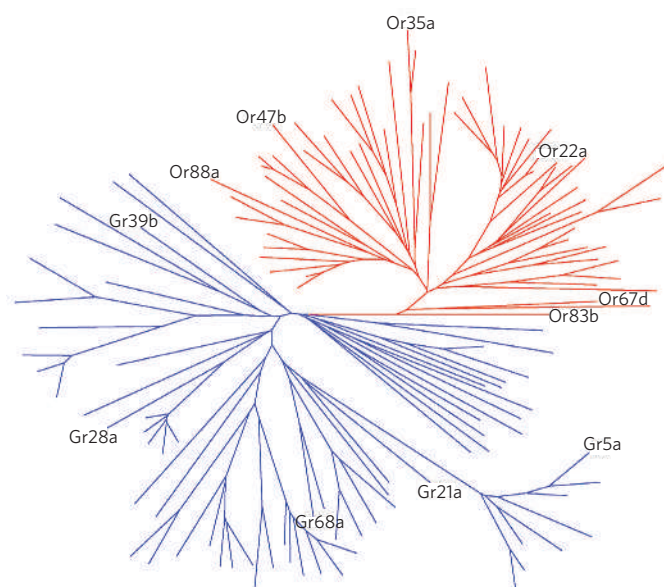


Figure 4 | Tree of *Or* (Odorant receptor) and *Gr* (Gustatory receptor) families of *Drosophila melanogaster*, which belong to a chemoreceptor superfamily. The length of each branch is a measure of sequence divergence. *Or* family is in red; *Gr* family is in blue. (Data from ref. 46.)

beans, alfalfa and cotton, and in the silk moth *B. mori*^{57,58}. Although most individual *Drosophila* odorant receptor genes have no direct *A. gambiae* orthologue, one odorant receptor gene, *Or83b*, is well conserved and is expressed in most ORNs^{59,60}. Functional analysis of *Or83b* has shown that it forms heterodimers with other odorant receptor proteins⁶¹ and that it is required for normal localization of odorant receptor proteins to the cell surface⁶².

Insect taste receptors were identified in 2000, again in *Drosophila*, through analysis of genomic DNA sequences⁶³. The *Drosophila* genome contains 60 gustatory receptor (*Gr*) genes⁴⁶ (Fig. 4), of which one, *Gr5a*, has been shown by mutational analysis to be essential for behavioural and physiological responses to the sugar trehalose⁶⁴. Although mutation of *Gr5a* reduced the response of sugar-sensitive cells to trehalose, it did not affect the response to sucrose, supporting a model in which individual taste neurons contain multiple receptors. Gustatory receptor genes are expressed in taste cells of the mouthparts and the legs^{63,65,66}; cells in the male foreleg that express *Gr68a* have been shown to be required for normal courtship behaviour, suggesting that *Gr68a* may detect a pheromone transferred by contact⁶⁷. A subset of gustatory receptor genes are expressed in cells that mediate aversive responses, consistent with a role for these genes in encoding receptors for bitter compounds^{68,69}. A family of gustatory receptor genes has also been identified in *A. gambiae*⁵⁵.

As detailed above, PBPs were first discovered in moths. They have subsequently been identified in insect taxa as diverse as scarab beetles and termites^{28,70}. Analysis of the *Drosophila* genome sequence has revealed a family of 51 orthologues⁷¹, known as odorant-binding proteins (OBPs). *A. gambiae* has 57 of these⁷².

New opportunities for insect pest control

The identification of molecular components that interact with odors and tastants offers new opportunities for pest control. First, the isolation of odorant receptors allows screening for compounds that strongly excite or inhibit them. Potent compounds could be useful as repellents or as attractants in traps. It may be possible, for example, to manipulate *Anopheles* receptors that detect human odours and help guide the mosquito toward a human host. Odorants that excite such receptors could be useful as trapping agents, and odorants that block them could be useful components of an insect repellent. Compounds that excite detectors of oviposition sites or food sources could also be effective trapping agents.

Useful odorants might include natural components of human sweat, oviposition sites or food sources, or other compounds that are not found in such sources but that act strongly on these receptors.

In addition to compounds that excite or inhibit odour receptors, other compounds might be useful for odour masking. Some odorants that do not elicit a response from an individual odour receptor have been shown to block responses normally elicited from the receptor by other compounds^{73–75}. Mosquito receptors that signal the presence of a human host would be especially interesting targets for odour-masking compounds.

Almost all insect odour receptors show remarkable sequence divergence across species, making them interesting targets from an ecological point of view. Some conventional insecticides inflict considerable collateral damage on a wide range of other species because they act on a highly conserved enzyme or ion channel. By contrast, the great sequence divergence among odour receptors suggests the possibility of identifying attractants that at low concentrations show more specificity for particular pests than do insecticides.

To identify compounds that excite, inhibit or block odour receptors, functional assays are needed. The empty neuron system developed in *Drosophila* has been used successfully as an *in vivo* expression system for odour receptors from an insect pest species: when the mosquito receptors AgOr1 and AgOr2 were expressed in this system, they responded to phenolic compounds, with different specificities⁷⁶. AgOr1, which mediated responses to the human sweat component 4-methylphenol, is a female-specific receptor³⁴, which is of special interest because only the female mosquitoes seek human hosts and bite. Compounds that potentially excite receptors such as AgOr1 could be effective in attracting mosquitoes to traps, and compounds that block such receptors could be useful as well. Heterologous expression in *Xenopus* oocytes has been used successfully not only with *Drosophila* odourant receptors⁴⁸ but also to show that two *B. mori* receptors respond to pheromones⁵⁸.

OBPs and PBPs might be useful targets for insect pest control, as suggested by the finding that mutation of a *Drosophila* OBP reduced response to the pheromone *cis*-vacccenyl acetate⁷⁷. Another interesting strategy is to interfere with enzymes that degrade pheromones or odorants⁷⁸. In addition, it may be feasible to impair specific receptor neurons important for host-seeking in many disease vectors, such as the neurons that sense carbon dioxide or ammonia, once their molecular receptors have been identified. Finally, it is possible that blocking the function of Or83b and its orthologues might affect the function of many odour receptors simultaneously⁵⁹.

Recent success in transforming the genomes of mosquitoes and other insect pests suggests it could be possible to manipulate chemosensory systems 'from the inside' — that is, through molecular genetic manipulation. For example, there has been interest in the possibility of converting mosquitoes from anthropophily to zoophily by manipulation of their odour receptors. However, the release of transgenic mosquitoes in the wild would be complicated by uncertainty as to their long-term fitness and the unpredictability of environmental consequences; moreover, there are political and social complications to the release of transgenic mosquitoes in areas of human habitation.

Another opportunity is to take advantage of the great progress made recently in insect chemosensation to identify the molecular mechanisms by which DEET and other effective agents repel insects. Identification of the targets and pathways through which such agents act would allow screening for alternative repellents.

A virtue of laboratory-based molecular, cellular and physiological screens is their convenience and rapidity. Collections of non-toxic, inexpensive, and stable compounds can be screened relatively quickly to identify candidates that might be useful in altering chemosensory-based behaviour. The rate at which compounds can be tested in the field remains low, and so a rapid means of identifying likely candidates would be of great benefit. There are important advantages to the identification of numerous compounds useful in pest management: for example, insects are less likely to develop resistance to a cocktail of agents than to a single agent.

However, an important requirement in maximizing the value of such screens is greater understanding of a problem that is important to basic neuroscience as well as to applied entomology: the logic that connects the responses of chemosensory neurons to behaviour. At present, there is no rational basis for predicting which compounds that affect chemosensory neurons in the laboratory will affect behavioral response in the field. Thus, there is a pressing need for understanding the processing of chemosensory information. Are the number of receptors excited by an odorant, or the summed strengths of their responses, useful parameters in predicting the efficiency with which the odorant attracts insects to a trap? Why do some odorants act synergistically? Do particular mosquito receptors activate a circuit that controls repulsion? Insight into the circuitry of insect chemosensation should be valuable in fully realizing the opportunities provided by progress in understanding of its peripheral basis.

Finally, we note that although it is difficult to predict the ultimate contributions of molecular- and cellular-based approaches to insect pest control, it is critical to maintain a perspective on the enormity of the problems that insect pests impose. For example, a new approach that reduced the incidence of malaria by as little as 1% could spare 5 million humans from debilitating disease. Thus, it will be of great interest to determine whether new chemosensory-based approaches can contribute, in conjunction with existing strategies, to the control of insects that do such immense damage to world health and agriculture. ■

1. Snow, R. W., Guerra, A. A., Noor, A. M., Myint, H. Y. & Hay, S. I. The global distribution of clinical episodes of *Plasmodium falciparum* malaria. *Nature* **434**, 214–217 (2005).
2. WHO & UNICEF *World Malaria Report 2005* [online] <<http://www.rbm.who.int/wmr2005>> (2005).
3. Smallegange, R. C., Qiu, Y. T., Van Loon, J. J. A. & Takken, W. Synergism between ammonia, lactic acid and carboxylic acids as kairomones in the host-seeking behaviour of the malaria mosquito *Anopheles gambiae sensu stricto* (Diptera: Culicidae). *Chem. Senses* **30**, 145–152 (2005).
4. Qiu, Y. T., Smallegange, R. C., van Loon, J. J. A. & Takken, W. Interindividual variation in odour-mediated attractiveness of human odours to the malaria mosquito *Anopheles gambiae*. *Med. Vet. Entomol.* **20**, 280–287 (2006).
5. Schreck, C. E., Kline, D. L. & Carlson, D. A. Mosquito attraction to substances from the skin of different humans. *J. Am. Mosquito Control Assoc.* **6**, 406–410 (1990).
6. White, G. B. *Anopheles gambiae* complex and disease transmission in Africa. *Trans. R. Soc. Trop. Med. Hyg.* **68**, 278–299 (1974).
7. Pates, H. V., Takken, W., Stuke, K. & Curtis, C. F. Differential behaviour of *Anopheles gambiae sensu stricto* (Diptera: Culicidae) to human and cow odours in the laboratory. *Bull. Entomol. Res.* **91**, 289–296 (2001).
8. Pimentel, D. Diversification of biological control strategies in agriculture. *Crop Protection* **10**, 243–253 (1991).
9. Haney, P. B. in *Boll Weevil Eradication in the United States through 1999* (eds Dickerson, W. A. et al.) 7–24 (The Cotton Foundation Publisher, Memphis, 2001).
10. Harris, K. F. & Maramorosch, K. (eds) *Vectors of Plant Pathogens* (Academic, New York, 1980).
11. Judd, J. G. R. & Borden, J. Distant olfactory response of the onion fly, *Delia antiqua*, to host-plant odour in the field. *Physiol. Entomol.* **14**, 429–441 (1989).
12. Zhang, A. et al. Identification of a new blend of apple volatiles attractive to the apple maggot, *Rhagoletis pomonella*. *J. Chem. Ecol.* **25**, 1221–1231 (1999).
13. Glendinning, J. I., Davis, A. & Ramaswamy, S. Contribution of different taste cells and signaling pathways to the discrimination of 'bitter' taste stimuli by an insect. *J. Neurosci.* **22**, 7281–7287 (2002).
14. Del Campo, M. L. et al. Host recognition by the tobacco hornworm is mediated by a host plant compound. *Nature* **411**, 186–189 (2001).
15. Fraenkel, G. S. The raison d'être of secondary plant substances. *Science* **129**, 1466–1470 (1959).
16. Metcalf, R. L., Metcalf, R. A. & Rhodes, A. M. Cucurbitacins as kairomones for diabrotic beetle beetles. *Proc. Natl Acad. Sci. USA* **77**, 3769–3772 (1980).
17. Witzgall, P., Lindblom, T., Bengtsson, M. & Tóth, M. The Pherolist Phero Net [online] <<http://www-pherolist.slu.se/pherolist.php>> (2004).
18. Mustaparta, H. in *Insect Pheromone Research: New Directions* (eds Cardé, R. T. & Minks, A. K.) 144–163 (Chapman & Hall, New York, 1997).
19. Butenandt, A., Beckmann, R., Stamm, D. & Hecker, E. Über den Sexual-Lockstoff des Seidenspinners *Bombyx mori*. Reindarstellung und Konstitution. *Z. Naturforsch.* **14**, 283–284 (1959).
20. Shields, V. D. C. & Hildebrand, J. G. Responses of a population of antennal olfactory receptor cells in the female moth *Manduca sexta* to plant-associated volatile organic compounds. *J. Comp. Physiol. A* **186**, 1135–1151 (2001).
21. De Bruyne, M., Clyne, P. J. & Carlson, J. R. Odor coding in a model olfactory organ: the *Drosophila* maxillary palp. *J. Neurosci.* **19**, 4520–4532 (1999).
22. Vogt, R. G. & Riddiford, L. M. Pheromone binding and inactivation by moth antennae. *Nature* **293**, 161–163 (1981).
23. Sandler, B. H., Nikonova, L., Leal, W. S. & Clardy, J. Sexual attraction in the silkworm moth: structure of the pheromone-binding protein-bombykol complex. *Chem. Biol.* **7**, 143–151 (2000).

24. Horst, R. *et al.* NMR structure reveals intramolecular regulation mechanism for pheromone binding and release. *Proc. Natl Acad. Sci. USA* **98**, 14374–14379 (2001).
25. Campanacci, V. *et al.* Revisiting the specificity of *Mamestra brassicae* and *Antheraea polyphemus* pheromone-binding proteins with a fluorescence binding assay. *J. Biol. Chem.* **276**, 20078–20084 (2001).
26. Maida, R., Krieger, J., Gebauer, T., Lange, U. & Ziegelberger, G. Three pheromone-binding proteins in olfactory sensilla of the two silkworm species *Antheraea polyphemus* and *Antheraea pernyi*. *Eur. J. Biochem.* **267**, 2899–2908 (2000).
27. Kaissling, K.-E. Olfactory perireceptor and receptor events in moths: a kinetic model. *Chem. Senses* **26**, 125–150 (2001).
28. Leal, W. S. Pheromone reception. *Top. Curr. Chem.* **240**, 1–36 (2004).
29. Schoonhoven, L. M., Van Loon, J. J. A. & Dicke, M. *Insect-Plant Biology* (Oxford Univ. Press, UK, Oxford, 2005).
30. Schoonhoven, L. M., Blaney, W. M. & Simmonds, M. S. J. in *Insect-Plant Interactions* (ed. Bernays, E. A.) 59–79 (CRC, Boca Raton, 1992).
31. Mitchell, B. K. Physiology of an ATP receptor in labellar sensilla of the tsetse fly *Glossina morsitans morsitans* Westw. (Diptera: Glossinidae). *J. Exp. Biol.* **65**, 259–271 (1976).
32. Van der Goes van Naters, W. M. & Rinkes, T. H. N. Taste stimuli for tsetse flies on the human skin. *Chem. Senses* **18**, 437–444 (1993).
33. Van der Goes van Naters, W. M. & Den Otter, C. J. Amino acids as taste stimuli for tsetse flies. *Physiol. Entomol.* **23**, 278–284 (1998).
34. Gilbert, I. H., Gouck, H. K. & Smith, C. N. New mosquito repellents. *J. Econ. Entomol.* **48**, 741–743 (1955).
35. Leak, S. G. A. *Tsetse Biology and Ecology: Their Role in the Epidemiology and Control of Trypanosomiasis* (CABI, Wallingford, 1999).
36. Kuhar, T. P., Mori, K. & Dickens, J. C. Potential of a synthetic aggregation pheromone for integrated pest management of Colorado potato beetle. *Agric. For. Entomol.* 77–81 (2006).
37. Arn, H. & Louis, F. in *Insect Pheromone Research: New Directions* (eds Cardé, R. T. & Minks, A. K.) 377–382 (Chapman & Hall, New York, 1997).
38. Schmutterer, H. Properties and potential of natural pesticides from the neem tree, *Azadirachta indica*. *Annu. Rev. Entomol.* **35**, 271–297 (1990).
39. Koul, O. *Insect Antifeedants* (CRC, Boca Raton, 2005).
40. Klun, J. A. *et al.* Comparative resistance of *Anopheles albimanus* and *Aedes aegypti* to N,N-diethyl-3-methylbenzamide (Deet) and 2-methylpiperidinyl-3-cyclohexen-1-carboxamide (AI3-37220) in laboratory human-volunteer repellent assays. *J. Med. Entomol.* **41**, 418–422 (2004).
41. Mota-Sanchez, D., Hollingworth, R. M., Grafius, E. J. & Moyer, D. D. Resistance and cross-resistance to neonicotinoid insecticides and spinosad in the Colorado potato beetle, *Leptinotarsa decemlineata* (Say) (Coleoptera: Chrysomelidae). *Pest Manag. Sci.* **62**, 30–37 (2006).
42. Glare, T. R. & O'Callaghan, M. *Bacillus thuringiensis: Biology, Ecology and Safety* (John Wiley & Sons, Chichester, 2000).
43. Hajek, A. *Natural Enemies: An Introduction to Biological Control* (Cambridge Univ. Press, Cambridge, UK, 2004).
44. Clyne, P. J. *et al.* A novel family of divergent seven-transmembrane proteins: candidate odorant receptors in *Drosophila*. *Neuron* **22**, 327–338 (1999).
45. Vosshall, L. B., Amrein, H., Morozov, P. S., Rzhetsky, A. & Axel, R. A spatial map of olfactory receptor expression in the *Drosophila* antenna. *Cell* **96**, 725–736 (1999).
46. Robertson, H. M., Warr, C. G. & Carlson, J. R. Molecular evolution of the insect chemoreceptor gene superfamily in *Drosophila melanogaster*. *Proc. Natl Acad. Sci. USA* **100**, 14537–14542 (2003).
47. Stortkuhl, K. F. & Kettler, R. Functional analysis of an olfactory receptor in *Drosophila melanogaster*. *Proc. Natl Acad. Sci. USA* **98**, 9381–9385 (2001).
48. Wetzel, C. H. *et al.* Functional expression and characterization of a *Drosophila* odorant receptor in a heterologous cell system. *Proc. Natl Acad. Sci. USA* **98**, 9377–9380 (2001).
49. Dobritsa, A. A., Van der Goes van Naters, W., Warr, C. G., Steinbrecht, R. A. & Carlson, J. R. Integrating the molecular and cellular basis of odor coding in the *Drosophila* antenna. *Neuron* **37**, 827–841 (2003).
50. Hallem, E. A., Ho, M. G. & Carlson, J. R. The molecular basis of odor coding in the *Drosophila* antenna. *Cell* **117**, 965–979 (2004).
51. Kreher, S. A., Kwon, J. Y. & Carlson, J. R. The molecular basis of odor coding in the *Drosophila* larva. *Neuron* **46**, 445–456 (2005).
52. Goldman, A. L., Van der Goes van Naters, W., Lessing, D., Warr, C. G. & Carlson, J. R. Coexpression of two functional odor receptors in one neuron. *Neuron* **45**, 661–666 (2005).
53. Hallem, E. A. & Carlson, J. R. Coding of odors by a receptor repertoire. *Cell* **125**, 143–160 (2006).
54. Fox, A. N., Pitts, R. J., Robertson, H. M., Carlson, J. R. & Zwiebel, L. J. Candidate odorant receptors from the malaria vector mosquito *Anopheles gambiae* and evidence of down-regulation in response to blood feeding. *Proc. Natl Acad. Sci. USA* **98**, 14693–14697 (2001).
55. Hill, C. A. *et al.* G protein-coupled receptors in *Anopheles gambiae*. *Science* **298**, 176–178 (2002).
56. Krieger, J. *et al.* A divergent gene family encoding candidate olfactory receptors of the moth *Heliothis virescens*. *Eur. J. Neurosci.* **16**, 619–628 (2002).
57. Sakurai, T. *et al.* Identification and functional characterization of a sex pheromone receptor in the silkworm *Bombyx mori*. *Proc. Natl Acad. Sci. USA* **101**, 16653–16658 (2004).
58. Nakagawa, T., Sakurai, T., Nishioka, T. & Touhara, K. Insect sex-pheromone signals mediated by specific combinations of olfactory receptors. *Science* **307**, 1638–1642 (2005).
59. Jones, W. D., Nguyen, T. A., Kloss, B., Lee, K. J. & Vosshall, L. B. Functional conservation of an insect odorant receptor gene across 250 million years of evolution. *Curr. Biol.* **15**, R119–R121 (2005).
60. Krieger, J., Klink, O., Mohl, C., Raming, K. & Breer, H. A candidate olfactory receptor subtype highly conserved across different insect orders. *J. Comp. Physiol. A* **189**, 519–526 (2003).
61. Neuhaus, E. M. *et al.* Odorant receptor heterodimerization in the olfactory system of *Drosophila melanogaster*. *Nature Neurosci.* **8**, 15–17 (2005).
62. Larsson, M. C. *et al.* Or83b encodes a broadly expressed odorant receptor essential for *Drosophila* olfaction. *Neuron* **43**, 703–714 (2004).
63. Clyne, P. J., Warr, C. G. & Carlson, J. R. Candidate taste receptors in *Drosophila*. *Science* **287**, 1830–1834 (2000).
64. Dahanukar, A., Foster, K., Van der Goes van Naters, W. M. & Carlson, J. R. A Gr receptor is required for response to the sugar trehalose in taste neurons of *Drosophila*. *Nature Neurosci.* **4**, 1182–1186 (2001).
65. Scott, K. *et al.* A chemosensory gene family encoding candidate gustatory and olfactory receptors in *Drosophila*. *Cell* **104**, 661–673 (2001).
66. Dunipace, L., Meister, S., McNealy, C. & Amrein, H. Spatially restricted expression of candidate taste receptors in the *Drosophila* gustatory system. *Curr. Biol.* **11**, 822–835 (2001).
67. Bray, S. & Amrein, H. A putative *Drosophila* pheromone receptor expressed in male-specific taste neurons is required for efficient courtship. *Neuron* **39**, 1019–1029 (2003).
68. Wang, Z., Singhvi, A., Kong, P. & Scott, K. Taste representations in the *Drosophila* brain. *Cell* **117**, 981–991 (2004).
69. Thorne, N., Chromey, C., Bray, S. & Amrein, H. Taste perception and coding in *Drosophila*. *Curr. Biol.* **14**, 1065–1079 (2004).
70. Vogt, R. G., Callahan, F. E., Rogers, M. E. & Dickens, J. C. Odorant binding protein diversity and distribution among the insect orders, as indicated by LAP, an OBP-related protein of the true bug *Lygus lineolaris* (Hemiptera, Heteroptera). *Chem. Senses* **24**, 481–495 (1999).
71. Hekmat-Scafe, D. S., Scafe, C. R., McKinney, A. J. & Tanouye, M. A. Genome-wide analysis of the odorant-binding protein gene family in *Drosophila melanogaster*. *Genome Res.* **12**, 1357–1369 (2002).
72. Xu, P. X., Zwiebel, L. J. & Smith, D. P. Identification of a distinct family of genes encoding atypical odorant-binding proteins in the malaria vector mosquito, *Anopheles gambiae*. *Insect Mol. Biol.* **12**, 549–560 (2003).
73. Araneda, R. C., Kini, A. D. & Firestein, S. The molecular receptive range of an odorant receptor. *Nature Neurosci.* **3**, 1248–1255 (2000).
74. Oka, Y., Omura, M., Kataoka, H. & Touhara, K. Olfactory receptor antagonism between odorants. *EMBO J.* **23**, 120–126 (2004).
75. Spehr, M. *et al.* Identification of a testicular odorant receptor mediating human sperm chemotaxis. *Science* **299**, 2054–2058 (2003).
76. Hallem, E. A., Fox, A. N., Zwiebel, L. J. & Carlson, J. R. Olfaction: mosquito receptor for human-sweat odorant. *Nature* **427**, 212–213 (2004).
77. Xu, P., Atkinson, R., Jones, D. N. & Smith, D. P. *Drosophila* OBP LUSH is required for activity of pheromone-sensitive neurons. *Neuron* **45**, 193–200 (2005).
78. Maibeche-Coisne, M., Nikonov, A. A., Ishida, Y., Jacquin-Joly, E. & Leal, W. S. Pheromone anosmia in a scarab beetle induced by *in vivo* inhibition of a pheromone-degrading enzyme. *Proc. Natl Acad. Sci. USA* **101**, 11459–11464 (2004).
79. De Bruyne, M., Foster, K. & Carlson, J. R. Odor coding in the *Drosophila* antenna. *Neuron* **30**, 537–552 (2001).

Acknowledgments Our research is funded by the NIH, a Senior Scholar Award from the Ellison Medical Foundation (J.R.C.) and a grant from the Foundation for the NIH through the Grand Challenges in Global Health initiative.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence should be addressed to J.R.C. (john.carlson@yale.edu).

Pheromonal communication in vertebrates

Peter A. Brennan¹ & Frank Zufall²

Recent insights have revolutionized our understanding of the importance of chemical signals in influencing vertebrate behaviour. Previously unknown families of pheromonal signals have been identified that are expanding the traditional definition of a pheromone. Although previously regarded as functioning independently, the main olfactory and vomeronasal systems have been found to have considerable overlap in terms of the chemosignals they detect and the effects that they mediate. Studies using gene-targeted mice have revealed an unexpected diversity of chemosensory systems and their underlying cellular and molecular mechanisms. Future developments could show how the functions of the different chemosensory systems are integrated to regulate innate and learned behavioural and physiological responses to pheromones.

The term pheromone was introduced by Karlson and Lüscher in 1959 (ref. 1). They defined pheromones as “substances secreted to the outside of an individual and received by a second individual of the same species in which they release a specific reaction, for example, a definite behaviour or developmental process”. Originally, two types of pheromonal signalling were recognized. Releaser pheromones elicit an immediate behavioural response — for example, 2-methylbut-2-enal, which is produced in rabbit milk. This single molecule elicits stereotyped nipple-search behaviour in rabbit pups, and is vital for them to locate the nipples during the brief daily period of suckling². By contrast, primer pheromones mediate more slowly developing and longer-lasting changes to endocrine state or development. For example, α -farnesene is one of several testosterone-dependent constituents of male mouse urine, which accelerate puberty of pre-pubertal female mice³.

As more examples of intraspecific chemical signalling have been identified, the original definition of a pheromone has appeared excessively rigid, and the term has frequently been used for any form of innate, intraspecific chemical communication⁴. Thus, many investigators also recognize the category of signaller pheromones. These convey information about the sender, such as individual or group identity, which are important for parent–offspring recognition and mate choice^{5,6}. A further category — modulator pheromones — has been proposed to affect mood and thought processes in humans, although this has yet to gain wide acceptance⁷.

This review seeks to highlight the diverse chemical nature of vertebrate pheromones and recent advances in our understanding of the chemosensory systems and neural mechanisms underlying such pheromonal effects. Although many of the principles discussed apply to a wide range of vertebrates, this review concentrates on mammalian pheromones, particularly those of rodents, in which most of the recent advances have occurred. These advances have revealed that both volatile molecules and non-volatile peptides and proteins can provide sex-specific and individual-specific cues. The same chemosignals can be detected by diverse chemosensory systems, which have access to the same behavioural and physiological outputs. However, although major advances have been made recently, our knowledge of the roles and mechanisms of pheromonal communication in vertebrates is still fragmentary. Future progress will require analysis at all levels, including the use of analytical chemistry to characterize the often complex mixture of chemical signals,

and molecular genetics to target specific receptor systems for phenotypic analysis at the cellular and behavioural levels.

The diversity of vertebrate pheromonal signals

Pheromones come in a wide variety of chemical forms. Their most significant features are their size and their polarity, which are the major factors determining their volatility in air and solubility in water, respectively. Thus, in the terrestrial environment, attractant and alarm pheromones, which by their nature act at a distance, are typically small and volatile. Methylthiomethanethiol (MTMT), which is present in male mouse urine and attracts investigation by females⁸, is a good example. By contrast, pheromones that convey information about specific individuals are likely to be relatively non-volatile — for instance, proteins or peptides — so that they do not disperse and can be more reliably associated with the producer. The situation differs in the aquatic environment, where solubility is the most important factor and even relatively high-molecular-weight peptides and proteins can have an attractant role. For example, the decapeptide sodefrin is produced by the abdominal gland of the male newt *Cynops pyrrhogaster*, and when released into water it attracts females of the same, but not closely related, species⁹. As discussed below, peptide and protein pheromones are also used widely in mammalian chemical communication^{10–15}. Other examples of aquatic attractant pheromones include the mixture of sulphated steroids and bile acids recently identified as migratory pheromones of lampreys¹⁶.

The structural constraints imposed by common metabolic pathways mean that similar molecules may be used as pheromones by different species. In these instances, the species specificity of the signal may be maintained by a pheromone blend or by the context in which it occurs. For example, the androgen derivatives 5 α -androst-16-en-3-one and 5 α -androst-16-en-3-ol, which are found in boar saliva, act synergistically as releaser pheromones to attract sows and elicit a receptive posture¹⁷. Similarly, although MTMT increases the attractiveness of male urine to females in the presence of other urinary cues, it is relatively ineffectual when presented on its own⁸. Male mouse urine also contains high concentrations of the testosterone-dependent volatiles (*R,R*)-3,4-dehydro-exo-brevicomin (DHB) and (*S*)-2-sec-butyl-4,5-dihydrothiazole (SBT). When sensed in the appropriate social context, or administered in urine from castrated males, these compounds act as releaser pheromones, eliciting aggression in male mice¹⁸. But pheromones do not always fit neatly into a single category and, along with their releaser effects on

¹Department of Physiology, University of Bristol, Medical School Building, University Walk, Bristol BS8 1TD, UK. ²Department of Anatomy and Neurobiology, University of Maryland, School of Medicine, 20 Penn Street, Baltimore, Maryland 21201, USA.

male aggression, DHB and SBT also have primer effects, accelerating puberty in prepubertal female mice and inducing and synchronizing oestrus in adults^{3,18}.

Multiple chemosensory systems mediate pheromonal effects

Mammals have several chemosensory systems capable of detecting pheromones (Fig. 1). The main olfactory epithelium (MOE) consists predominantly of ciliated olfactory sensory neurons (OSNs), which project to the main olfactory bulb (MOB). Each of these expresses a single receptor type from a large family — mice have almost 1,300 — of olfactory receptors¹⁹. Recently, a second family of G-protein-coupled receptors, the trace-amine-associated receptors (TAARs), which recognize volatile amines present in urine, has also been shown to be expressed in OSNs²⁰.

In addition to the MOE, most mammals have a vomeronasal organ (VNO)^{21,22}. This is a blind-ended tube located in the nasal septum, and contains vomeronasal sensory neurons (VSNs) with microvillar morphology, which project to the accessory olfactory bulb (AOB). Two classes of vomeronasal receptor have been identified, the V1rs and the V2rs (ref. 22). They belong to the seven-transmembrane-domain G-protein-coupled receptor superfamily, but share little homology with each other or with the olfactory receptors, suggesting that they have distant evolutionary origins. Gene targeting of a cluster of V1rs has established that these molecules are involved in responses to small organic pheromones such as 6-hydroxy-6-methyl-3-heptanone²³, which accelerates puberty in female mice¹⁸. No such information is yet available for V2rs.

In addition to their distinct morphology, VSNs have a different transduction mechanism from OSNs, involving a diacylglycerol-activated cation channel, which depends in part on the transient-receptor-potential channel 2 (*Trpc2*) gene²⁴. Analysis of the mouse genome has identified 137 functional receptors of the V1r class²⁵ as well as 60 or so potentially functional V2rs²⁶. Although a similar set-up has been found in other rodents, and in opossums, it is not typical of all mammals, let alone all vertebrates. Therefore, as with most aspects of pheromonal signalling, species differences must be borne in mind.

Notable evolutionary changes have occurred to chemosensory systems in association with the transition from aquatic to terrestrial environments. Fish do not have a VNO. Instead, they have a single olfactory organ containing a mixture of ciliated sensory cells expressing

olfactory-receptor-like receptors and microvillar cells expressing both V2r-like²⁷ and V1r-like²⁸ receptors. These microvillar sensory neurons seem to have segregated into a separate organ in early terrestrial vertebrates, at about the same time that the MOE was adapting to detect airborne odorants. However, the detailed story is likely to be complex²⁹, and the division between cell types is not absolute. Although most sensory cells in the mammalian MOE are ciliated and express olfactory receptors, also present are microvillar cells that seem to form a distinct chemosensory system³⁰.

The different locations and anatomical structures of the MOE and VNO have consequences for stimulus access. Whereas the MOE has access to stimuli in the nasal airstream, the VNO is connected to the nasal cavity by a narrow duct²¹. Stimuli are thought to gain access to the VNO by a vascular pumping mechanism that is activated in arousing situations³¹. This has led to the idea that the main olfactory system is specialized for detecting volatile, airborne molecules, whereas the vomeronasal system is specialized for the detection of non-volatile chemosignals, such as those in urine, skin, scent glands and reproductive secretions, after direct contact with the stimulus source. However, recent findings indicate that this functional division is not as absolute as was once thought.

VSNs can detect small organic molecules present in urine — for example, the male pheromones SBT and DHB^{32–34} — that bind to carrier proteins such as major urinary protein³⁵ (MUP; see below). *In vivo*, such molecules probably gain access to the VNO after direct contact with the stimulus source, which is consistent with increases in neuronal activity in the AOB after direct contact with an anaesthetized conspecific³⁶. There is some debate about whether the VNO can detect volatiles in the absence of physical contact. Katz's group failed to find neuronal responses in the AOB to volatile odorants or pheromones presented on cotton swabs without direct contact. However, relatively few cells were tested, and it is not clear whether recordings were made from the anterior subregion of the AOB, which receives projections from the volatile-sensing V1r class of VSNs²². By contrast, functional magnetic resonance imaging (fMRI) in anaesthetized mice showed robust changes in the activity of the AOB in response to urine odours delivered via the airstream³⁷. These responses were predominantly localized to the anterior part of the AOB, which is consistent with the projection of the volatile-sensing VSNs to the anterior subregion. Finally, even the seemingly reasonable hypothesis that the MOE responds only to airborne molecules will have to be rethought, as non-volatile stimuli have been shown to reach the MOE after direct contact with a stimulus¹⁰.

Mice with VNO ablation show a number of behavioural and physiological deficits in their response to chemosensory cues. For instance, male urinary chemosignals fail to elicit aggression in males with VNO lesions, and such males do not show the normal rise in luteinizing hormone (LH) levels in response to female chemosignals³⁸. Genetic ablation of *Trpc2* markedly reduces VSN responses and leads to specific deficits in male–male and maternal aggression^{39,40}. However, its effects on male sexual behaviour differ from those of physical ablation, which may be explained in part by the recent finding that some peptide-sensing VSNs are TRPC2-independent⁴¹ (Fig. 2).

These findings have focused attention on the VNO as a conveyer of pheromonal responses. But it has long been known that many pheromonal responses are mediated by the main olfactory system. For instance, the nipple-search pheromone that guides the suckling of rabbit pups is still effective after VNO lesions⁴². Similarly, in sows, the effects of androstenone and androstenol on sexual behaviour are unaffected by VNO lesions¹⁷. Furthermore, male mouse urinary attractant MTMT selectively excites neurons in the MOB, although whether its attractant effects are mediated by the main olfactory pathway remains to be seen⁸.

Genetic ablation of the cyclic-nucleotide-gated channel subunit CNGA2, which is vital for transduction in most mouse OSNs, causes behavioural deficits equally as marked as VNO lesions, including lack of chemosensory investigation of conspecifics and impaired mating⁴³. This does not necessarily mean that all of the effects seen in *Cnga2*-knockout

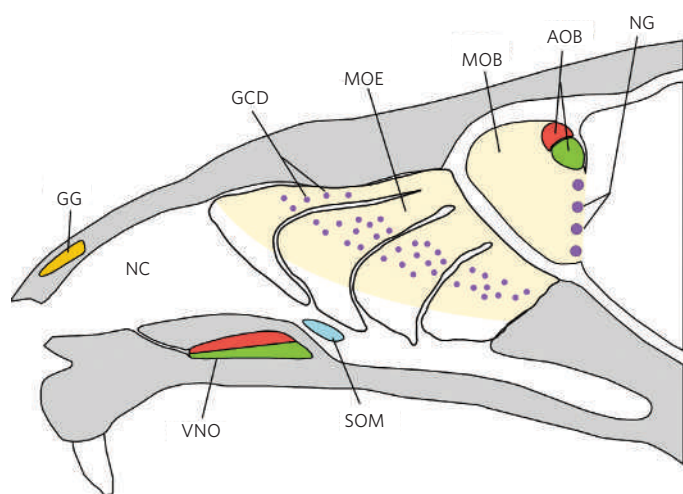


Figure 1 | The mouse olfactory system. Scheme highlighting the heterogeneity of chemosensory subsystems in the nose of the mouse (sagittal view). Sensory cells in each subsystem express specific signal detection molecules (not shown), which enables selective targeting of these populations^{22,45,46}. Chemosensory functions have not yet been established for the guanylyl cyclase type-D system (GCD) and Gruenberg ganglion (GG). NC, nasal cavity; NG, necklace glomeruli; SOM, septal organ of Masera.

animals are mediated solely by the main olfactory system. Main olfactory signals have an important role in attracting attention to stimuli to be investigated by the vomeronasal system⁴⁴, and possibly in activation of the vomeronasal pump. If this is the case, then some of the deficits seen in *Cnga2*-knockout mice may result from indirect effects on VNO function. Indeed, the results of genetic-ablation experiments must be interpreted with care, as they can be confounded by the function of other olfactory subsystems with as yet unknown roles in pheromone detection (Fig. 1). For instance, a subpopulation of MOE sensory neurons expresses a guanylyl cyclase-D signalling system rather than the canonical cyclic AMP–CNGA2 system⁴⁵. This chemosensory subsystem is not affected in *Cnga2*-knockout mice. There are further chemosensory systems about which little is known at present. The septal organ of Masera is an anatomically isolated area of main olfactory-like epithelium, situated ventrally to the MOE, and, although it projects to the ventral region of the MOB, its function is unclear⁴⁶. The Grueneberg ganglion, which is situated at the anterior end of the nasal cavity, is also likely to have a chemosensory role — a prediction based on its expression of olfactory marker protein and its projections to glomeruli in the dorsal region of the MOB⁴⁷.

MHC-associated cues as individuality signals

It has long been recognized that behaviours such as mate choice and parent–offspring interactions can be influenced by major histocompatibility (MHC) genotype. MHC molecules are encoded by a highly polymorphic family of genes that determine immunological identity at the tissue level. Hence, laboratory mice are more likely to mate with individuals of dissimilar MHC genotype from their own⁴⁸. Such disassortative mating is found in semi-natural conditions in which fewer MHC-homozygous offspring are born than would be expected from random matings⁶. The maternal recognition of young is also affected by MHC genotype. Maternal mice preferentially retrieve pups of the same MHC type as themselves⁵. Similarly, in a Y-maze test, mouse pups were found to prefer nest odours of maternal and sibling MHC type to unfamiliar MHC type⁵. Furthermore, mice can be trained to discriminate between urine odours of congenic mice that differ genetically only at their MHC H2 locus⁴⁹, implying that MHC genotype influences the volatile constituents of urine.

MHC class I proteins bind a vast range of nine-amino-acid peptides that result from proteasomal degradation of endogenous and foreign proteins, and present them at the cell surface for immune surveillance. The ability of mice to discriminate urine odours of MHC-congenic strains is consistent with the polymorphism in the peptide-binding groove of the MHC protein⁵⁰. Fragments of MHC class I proteins have been found in mouse urine, albeit at low levels⁵¹. But how could they influence urine odour? According to the carrier hypothesis, the peptide-binding groove releases its peptide ligand when the MHC protein is cleaved from the cell, and becomes available for binding small, volatile molecules⁵². Polymorphic variations in the peptide-binding-groove structure would determine the composition of volatiles bound by MHC class I proteins, and therefore influence body odour⁵⁰. Although there is no direct evidence for the hypothesis that MHC class I proteins are able to bind volatiles, MHC genotype has been shown to influence the profile of volatile fatty acids in urine⁵³. However, it should be noted that although the volatile profile affects the discriminability of mouse urine, MHC-dependent urinary volatiles have yet to be shown to convey individuality in a natural context.

More recent evidence has demonstrated that the MHC peptide ligands themselves can function as chemosignals, forming a direct link between individuality at the immunological and behavioural levels^{11,12}. VSNs of the V2r class respond highly sensitively to these nine-amino-acid peptides¹² (Fig. 2b). Both VSN responses and the specificity of peptide binding to MHC class I proteins are determined by the position of what are known as anchor residues. These are amino acids with bulky side chains that fit into pockets in the peptide-binding cleft. A particular MHC class I molecule can bind a variety of peptides that differ in amino-acid composition but have anchor residues constrained to specific positions along the peptide chain. The identities and positions of the anchor

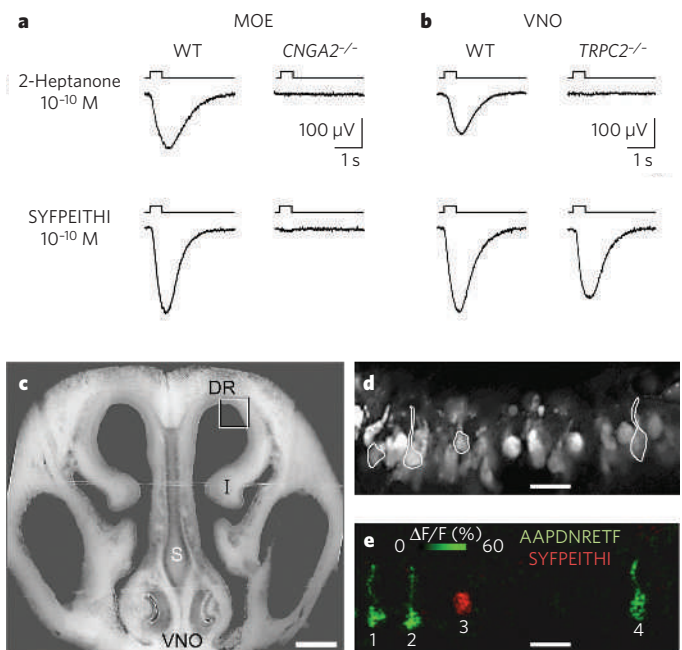


Figure 2 | Both the MOE and VNO detect pheromonal ligands at low concentrations. **a, b**, Electrophysiological recordings from wild-type (WT) and gene-targeted mice showing examples of field-potential responses to the same pheromone ligands at low concentrations (10⁻¹⁰ M) in the MOE and VNO. 2-Heptanone is a urinary constituent that has a primer pheromonal effect, extending the length of female oestrous cycles¹⁸. SYFPEITHI is a prototypical MHC class-I peptide ligand (for the H-2^d haplotype of BALB/c mice)¹². Note that both MOE responses are absent in mice with a targeted mutation in the cyclic-nucleotide-gated channel subunit CNGA2. The response to 2-heptanone is absent in the VNO of mice with a targeted mutation in the transient-receptor-potential channel TRPC2; however, the response to SYFPEITHI is relatively normal in these mutants. **c–e**, Large-scale confocal Ca²⁺ imaging with cellular resolution using a nasal coronal tissue slice identifies discrete populations of peptide-detecting olfactory sensory neurons in the mouse MOE. AAPPDNRETF is a prototypical MHC class-I peptide ligand for the H-2^b haplotype of C57BL/6 mice. A portion of the olfactory epithelium delimited by the white box in **c** is shown at higher magnification in **d** and **e**. **d**, Confocal fluorescence image of the fluo-4-loaded olfactory epithelium acquired at rest (grayscale). **e**, Pseudocolour image of the relative increase in peptide-induced Ca²⁺ fluorescence, showing four responsive OSNs. S, septal wall; DR, dorsal recess; I, endoturbinates I. (Images **a**, **c**, **d** and **e** reproduced, with permission, from ref. 10. Image **b** is based on data from ref. 41.)

residues of an MHC peptide ligand therefore convey information about the MHC class I protein that binds it. As the VSN responses show very similar dependence on anchor-residue position to those of the MHC class I proteins, they convey the MHC identity of the peptides. Indeed, the addition of synthetically produced MHC peptide ligands has been shown to influence the perceived strain identity of mouse urine, which determines mate recognition in the context of the pregnancy-block effect¹² (see below).

Neither the responses of the V2r-expressing VSNs to MHC peptide ligands (Fig. 2b) nor the effectiveness of these ligands in conveying individuality in the pregnancy-block effect require a functional TRPC2 channel⁴¹. This implies that the peptide-sensitive VSNs rely on a different transduction mechanism from the TRPC2-dependent transduction used by the V1r class of VSNs. The V2r class of vomeronasal receptors are structurally very different from the V1rs. V2rs have a large extracellular amino-terminal domain²², which might be involved in binding MHC peptide ligands. Intriguingly, V2rs are co-expressed with non-classical MHC Ib proteins^{54,55}, with which they might form a receptor complex⁵⁴. There are nine members of this non-classical MHC Ib family. However, their binding groove is unlikely to bind peptides, and their

role in MHC peptide recognition is unclear⁵⁶. Certain combinations of MHC Ib proteins are expressed with particular V2rs⁵⁵, and it is tempting to speculate that they might convey innate differences in vomeronasal signalling between different MHC types.

Highly specific responses to MHC peptide ligands have recently been recorded from the mouse MOE, although their dependence on anchor residues differs from the responses of VSNs¹⁰ (Fig. 2a, e). Hence, MHC-dependent signals can be sensed by both the main olfactory system and the vomeronasal system. MHC peptide ligands have also been shown to influence preferences of female sticklebacks in their choice of mate, in a manner consistent with the generation of optimum MHC diversity in their offspring⁵⁷. Given the ubiquity of class I MHC proteins, the role of their peptide ligands in signalling individuality may be widespread among vertebrates.

Major urinary proteins and territorial behaviour in mice

A classic example of the use of signalling pheromones is scent marking. Individuals deposit scent cues around their environment as long-lasting advertisements of their presence to competitors or potential mates. This is perhaps best understood in mice, in which dominant males 'paint' small urine marks throughout their territory, especially at the boundaries and along major access routes⁴⁴. Mice excrete large amounts of protein in their urine, 99% of which are MUPs. These are 18–20 kDa members of the lipocalin family of ligand-binding proteins, which are thought to have a chemosensory signalling role in a variety of species and behavioural contexts⁵⁸. MUPs bind volatile ligands including the testosterone-dependent male mouse pheromones DHB, SBT, (*E,E*)- α -farnesene, (*E*)- β -farnesene and 6-hydroxy-6-methyl-3-heptanone³. Moreover, the concentration of MUPs is four to five times higher in male than female mouse urine, and some MUP variants are found only in males⁵⁹. The release of volatiles from urine marks is prolonged by their binding to MUPs and serves to attract direct investigation of the non-volatile chemosignals. As the ligands are released over a period of time, they also convey the freshness of urine marks, which allows a male to advertise his territorial dominance and competitive ability to potential mates, without the potentially damaging consequences of direct confrontation with competitors⁴⁴.

Male mice countermark urine marks that have been left by an intruder, a behaviour that is driven by non-volatile components of urine (presumably MUPs) and is likely to be mediated by the VNO⁴⁴. An individual mouse in the wild produces a specific pattern of 5–15 variants from the polymorphic MUP family⁶⁰, and the rate of countermarking of urine marks is associated with their MUP profile¹³. The differences in amino-acid residues among MUP variants are largely localized to the surface of the protein, and, although some are in the ligand-binding calyx, they have so far been found to have limited influence on the relative binding of their volatile ligands³⁵. It remains to be determined whether MUPs interact directly with receptors, although recent evidence suggests that MUPs that have been stripped of their ligands have pheromonal activity in influencing ovulation¹⁴.

Neural coding of pheromonal information

The vomeronasal system is often thought of as a labelled-line system: a high degree of receptor-cell stimulus specificity is maintained in their neural projections to the central structures that control the behavioural or neuroendocrine output. Certainly, V1r-expressing VSNs seem to be good candidates for such a system, as they respond highly selectively. Moreover, their specificity does not broaden as the stimulus concentration is increased, as is typically found in OSNs³². However, the responses to MHC peptide ligands of V2r VSNs is likely to be considerably more complicated than a 'one ligand–one receptor–one cell' coding strategy⁶¹. Individual VSNs are frequently found to respond to MHC peptide ligands from more than one MHC type¹². It remains to be established whether the coding of MHC individuality by VSNs relies on a pattern recognition system, or whether the responses of individual VSNs to particular combinations of MHC peptides are representative of specific combinations of MHC alleles in a labelled-line manner.

Neurons in the AOB respond to specific strain and sex combinations during direct contact with the head and anogenital regions of an anaesthetized conspecific³⁶. This is interesting, because the V2r MHC-peptide sensitive VSNs that respond to strain identity project to the posterior subregion of the AOB, whereas the V1r VSNs that respond to testosterone-dependent urinary volatiles, such as SBT and DHB, project to the anterior subregion. However, it is likely that little integration of information occurs between these information streams at the level of the AOB⁶². So, is there another source of male-specific chemosignals that could be sensed by the V2rs? One possibility is the recently identified 7 kDa protein produced by the extraorbital lacrimal gland of male mice. This is encoded by a member of a family of at least 23 related genes that are expressed by other glands in the head region, including the salivary and Harderian glands, although their behavioural effects are currently unknown¹⁵.

The main olfactory system is thought to code information using a pattern-recognition system in which chemosensory specificity arises out of comparisons among input from receptor types with relatively broad and overlapping molecular-receptive ranges. Hence, the relative proportions of the MHC-dependent volatile constituents of urine odour are represented by broad and overlapping but statistically different patterns of activity in the MOB⁵³. However, there is evidence that OSN responses can be as selective as those seen in VSNs. For example, individual OSNs in the MOE have been found to respond to MHC peptide ligands from only a single MHC type¹⁰, and recordings from the MOB have shown highly selective responses to the male attractant pheromone MTMT at naturally occurring concentrations⁸.

Perhaps one of the most significant recent advances is the growth of appreciation of the complementary roles of the main olfactory and vomeronasal systems. OSNs and VSNs not only respond to overlapping sets of social chemosignals, but do so with extraordinarily high sensitivity (Fig. 2). The mouse MOE responds robustly to urinary volatiles, such as 2-heptanone, at concentrations of 10^{-10} M (ref. 10), which is comparable to the sensitivity of V1rb2-expressing VSNs³⁴. The responses of the two systems to MHC peptide ligands are also highly sensitive, with robust responses at 10^{-10} M for OSNs¹⁰ and 10^{-12} M for VSNs¹². Moreover, both the main olfactory system and the vomeronasal system have access to neural systems controlling endocrine and behavioural responses. For example, the vomeronasal pathway has long been recognized as projecting to the region of the anterior hypothalamus, where luteinizing-hormone releasing hormone (LHRH) neurons are located. A recent study using a barley lectin transsynaptic tracer expressed specifically in LHRH neurons has shown that they receive direct synaptic connections from the vomeronasal pathway⁶³. Moreover, this study and another using a viral transsynaptic tracer targeted to LHRH neurons have both shown abundant inputs from areas handling main olfactory information^{63,64}.

Pheromones and learning

Pheromones are generally thought to elicit innate responses that occur in naive animals without prior learning. However, one of the best-known examples of pheromonal effects in vertebrates is the selective pregnancy block (or Bruce effect)²¹, which is associated with the formation and maintenance of an olfactory recognition memory by the vomeronasal system^{21,41,65}. A female mouse learns the individual identity of the male's urinary chemosignals during mating. This is vital to prevent implantation failure and abortion of her mate's offspring upon subsequent exposure to his MHC peptide ligand-associated chemosignals¹² (Fig. 3). Learning depends on a mating-induced noradrenaline release in the AOB, which results in inhibition of her mate's pheromonal signal at the level of the AOB⁶⁵. This is proposed to disrupt transmission of the pregnancy-blocking signal, and results in decreased activation of central brain areas by her mate's pheromones^{66,67}, preventing the activation of the neuroendocrine mechanisms leading to pregnancy failure (Fig. 3).

Learning also has an important role in governing social odour preferences in mice. The preference of mouse pups for maternal odours is

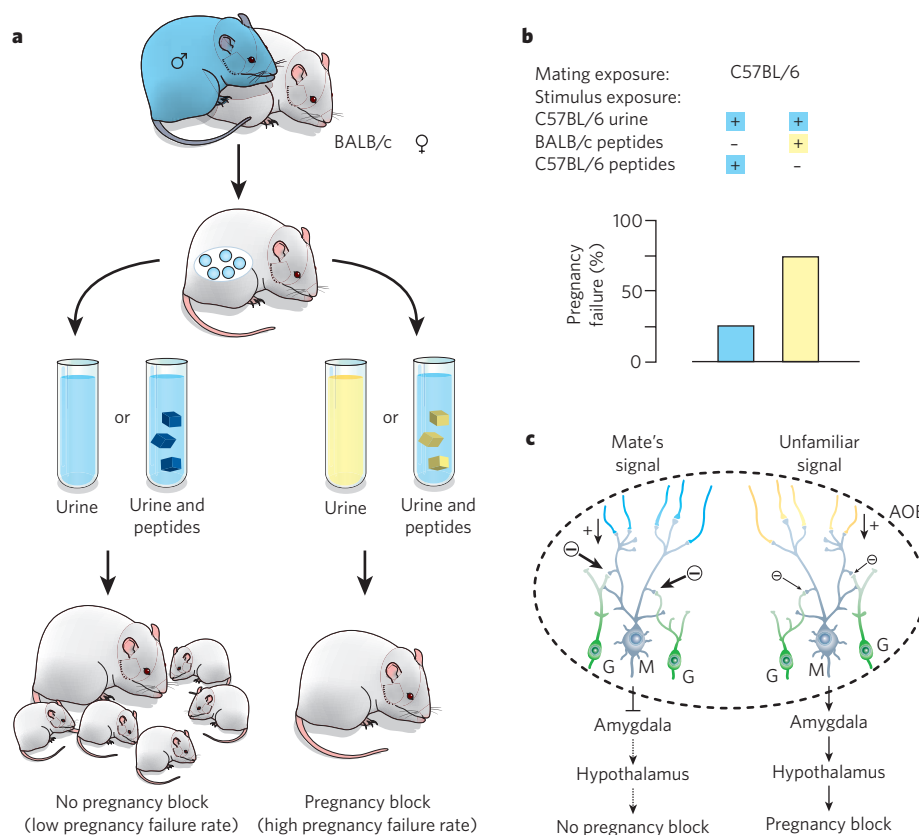


Figure 3 | The Bruce effect, one of the best-known examples of olfactory imprinting in adult vertebrates. Urine from unfamiliar males (yellow) — or familiar urine (blue) supplemented with unfamiliar (disparate) MHC class-1 peptide ligands — activates neuroendocrine mechanisms leading to pregnancy failure (**a**, **b**) by means of the AOB, amygdala and medial hypothalamus (**c**). The mated female learns to recognize the pheromones of the mating male during a sensitive period around mating. Subsequent exposure to her mate's urine activates mitral cells (M) in the AOB that are subject to enhanced inhibition from granule cell (G) interneurons. This increased inhibition is thought to disrupt the transmission of the mate's pheromonal signal to the amygdala, preventing pregnancy block. (Images **a** and **b** are based on data from ref. 12. Image **c** is based on data from refs 65–67, 99.)

influenced by neonatal learning in the nest environment⁵. Such neonatal learning of maternal and sibling odours can have lasting effects on MHC-associated mate preferences in adulthood, which can be partly reversed by cross-fostering to a mother of dissimilar MHC type^{68,69}. An intriguing question is the extent to which MHC-related social preferences are innate, and whether these pheromonal effects are mediated by MHC peptide ligands rather than by the MHC-dependent profile of volatiles that is thought to constitute individual body odour.

Learning is an important component of many other pheromonal responses. For example, the pheromone 2-methylbut-2-enal, found in rabbit milk, elicits highly stereotyped nipple-search behaviour in rabbit pups through the main olfactory system. However, if the doe's ventrum has been painted with an artificial odour, such as a perfume, then the pups learn, in a single exposure, to associate that odour with the pheromonal response⁷⁰. Subsequent presentation of the conditioned odour is sufficient to elicit full nipple-search behaviour in the absence of the nipple-search pheromone. Such an association of chemosensory cues and information from other sensory modalities is likely to play an important part in reinforcing pheromonal responses in natural contexts.

Vomeran stimulation seems to have intrinsic reinforcing and rewarding properties that not only maintain interest in social investigation but could also act as an unconditioned stimulus in the associative learning of main olfactory cues⁷¹. It is thought that odours conveyed by the main olfactory system become conditioned by vomeronasal stimuli and subsequently become sufficient to drive the same behavioural response in the absence of vomeronasal stimulation. For instance, VNO lesions severely disrupt mating in sexually naive hamsters, whereas VNO lesions in sexually experienced hamsters are less effective, as learned chemosensory cues sensed through the main olfactory system have become associated with mating⁷².

The brain's dopaminergic reward system is likely to be heavily involved in odour conditioning because both ejaculation and exposure to chemosignals from a receptive female are effective in increasing dopamine release in the nucleus accumbens of male rats⁷³. The increased activity

of neurons in the nucleus accumbens in response to a conditioned odour to which the animal was exposed during copulation is mediated through a main olfactory neural pathway⁷⁴, in contrast to the vomeronasal areas activated by chemosignals from a female in oestrus⁷⁵.

The posterior medial amygdala responds to socially relevant conspecific but not heterospecific stimuli⁷⁶. Moreover, homeodomain transcription factors delineate separate networks of neurons in the medial amygdala that are activated by reproductive or defensive chemosensory stimuli and project to hypothalamic areas controlling the appropriate behavioural responses⁷⁷. The cortical and medial regions of the amygdala receive direct input from the AOB and MOB, and form a hub in the networks governing social behaviour in mammals (Fig. 4). These regions have been implicated in social-recognition learning in rodents⁷⁸ and infant recognition in sheep⁷⁹.

Do human pheromones exist?

Can human pheromones be identified? What are their effects? And can they be exploited commercially? These are undoubtedly questions of great public interest, and a controversial area of research. This is partly because human social systems are too complex for human behaviour to be dominated by the drastic releaser effects of pheromones seen in other species. However, this does not mean that pheromonal signals did not influence the behaviour and physiology of our recent human ancestors, nor that they do not have subtle influences on humans today.

One of the main candidates for a source of human pheromones is the apocrine-gland secretions of the axilla, which, when subject to microbial action, produce the complex mixture of odorants responsible for body odour, including androgen derivatives and volatile acids⁸⁰. One of the major constituents of axillary secretions is (*E*)-3-methyl-2-hexanoic acid⁸¹. This is bound to apolipoprotein D, a member of the lipocalin family of ligand-binding proteins⁸², which are important transporters of pheromones in other species⁵⁸. However, although apocrine secretions are a likely source of potential human pheromones, little is known about the chemosensory systems involved in their detection. The VNO

is present early in human fetal development and the vomeronasal nerves have a vital role in guiding the migration of developing LHRH neurons to the hypothalamus⁸³. However, the epithelial diverticulum in the adult human nasal septum, referred to as the VNO, does not have a similar structure or function to the rodent VNO^{7,84}. Moreover, in humans, the gene encoding the TRPC2 cation channel is a pseudogene, having been impaired and removed from functional constraints about 23 million years ago, shortly before the separation of hominoids and Old World monkeys^{85,86}. This had been seen as the final nail in the coffin for the human VNO. However, the recent finding that V2r-type peptide-responding VSNs do not require TRPC2 channel function⁴¹ reminds us that our understanding of vomeronasal signal transduction is far from complete and that this evidence is not as conclusive as originally thought. More problematic for advocates of the existence of a human VNO are the consistent failure to find vomeronasal nerves projecting to the brain and the failure to find an AOB in adult humans⁸³.

The main olfactory system is therefore the most likely pathway for conveying any pheromonal effects in humans. Analysis of the human genome has revealed that most V1r-like genes are pseudogenes⁸⁷. Intriguingly, however, four functional human V1r-like genes have been found⁸⁸. One of these, V1RL1, is expressed in the olfactory mucosa and may well have a receptor function, although this does not necessarily mean that it responds to a pheromone⁸⁹. Indeed, members of the olfactory-receptor-gene family, the newly discovered TAAR family²⁰ or other as yet unknown receptors could function as pheromone receptors in humans as well as other species.

There is relatively good evidence for primer pheromonal effects in humans. Exposure to axillary secretions from females in the late follicular and ovulatory phases of their menstrual cycle have been found to shorten and lengthen, respectively, the cycles of recipient females⁹⁰. Furthermore, male axillary extracts seem to influence female reproductive state by affecting the frequency of LH release⁹¹. Evidence for signalling pheromones in humans is also reasonably strong. Several studies have found that females prefer odours from individuals of dissimilar MHC type^{92,93}, and this might even affect their choice of sexual partner⁹⁴. It remains to be determined whether MHC-peptides are detected by the MOE in humans and, if so, whether this has a similar role to its mediation of individual odour preferences in mice¹⁰.

The breast odour of human mothers has been reported to attract the attention of newborn babies and elicit movement towards the odour⁹⁵, which is comparable with the releaser effects of pheromones in other vertebrate species. However, whether pheromones mediate sexual attraction in adult humans is a much more complex issue⁷. Although there are reports of chemosensory effects on attractiveness ratings under laboratory conditions, there have not been many rigorous, placebo-controlled, double-blind studies on sexual activity in natural social situations. Those that have been performed must be interpreted with caution owing to the large differences between individuals in their opportunities for, and the nature of, their sexual interactions. The situation is further complicated by fluctuations in female olfactory sensitivity during the menstrual cycle, and reports that pheromones can have a modulatory effect in humans to alter their mood⁹⁶. This poses the question of whether changes in measures of sexual attraction or behaviour are actually secondary to changes in mood or general motivation. Functional MRI of brain activity provides a promising technique with which to study the neural basis of putative pheromonal effects in humans⁹⁷.

Future prospects

Recent developments have led to an appreciation of the diversity of chemosensory systems and their complementary roles in influencing vertebrate physiology and behaviour. However, our understanding of the nature of pheromones and the neural mechanisms underlying their effects is still fragmentary. Future developments in this fast-moving field will increasingly require researchers to integrate gene-targeting approaches with physiological and behavioural characterization. A greater understanding of the transduction mechanisms involved should allow selective and inducible ablation of specific receptor

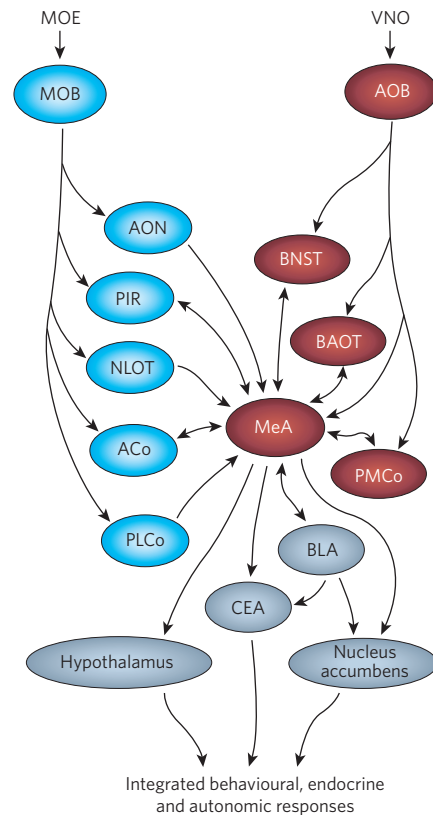


Figure 4 | The medial amygdala is an important hub that controls social behaviour. The medial amygdala (MeA) receives input from both the vomeronasal structures (red) and main olfactory structures (blue), and is a major site for the integration of chemosensory information with other sensory cues and hormonal states. The MeA elicits behavioural, endocrine and autonomic responses to pheromones through its output to the hypothalamus and central nucleus of the amygdala (CEA). Projections to the basolateral nuclei of the amygdala (BLA) and nucleus accumbens are likely to be involved in learned behavioural responses to conditioned stimuli. Interconnections of the individual structures not involving the MeA have been omitted for clarity. The central projections and possible integration sites from the septal organ of Maser, the Grueneberg ganglion and the GCD-cell system are unclear. ACo, anterior cortical amygdala; AON, anterior olfactory nucleus; BAOT, bed nucleus of the accessory olfactory tract; BNST, bed nucleus of the stria terminalis; NLOT, nucleus of the lateral olfactory tract; PIR, piriform cortex; PLCo, posterior lateral cortical amygdala; PMCo, posterior medial cortical amygdala. This simplified version of the connectivity of the MeA is based on neuroanatomical data from rats¹⁰⁰.

systems. However, the phenotypes produced are likely to be complicated by the interdependence and redundancy of the chemosensory systems and the important role of learning. An understanding of the behavioural effects needs to be coupled with gene-targeted tracing of specific receptor systems using transynaptic tracers to map their neural projections. Developments in imaging, telemetry and multiple-channel electrophysiological recording techniques promise to allow investigations of these neural systems to be conducted in semi-natural behavioural situations rather than an artificial laboratory environment.

The revolution in our understanding of vertebrate pheromonal communication has been, and will continue to be, driven by molecular-genetic approaches. However, an important limitation is that our understanding is based mainly on work in rodents, particularly mice. In some respects this is highly appropriate, given the importance of chemosensory information to them. However, pheromones are, by definition, species-specific signals and caution must be exercised in extrapolating between species, and between sexes, that have different reproductive strategies and behavioural priorities⁹⁸.

1. Karlson, P. & Lüscher, M. Pheromones: a new term for a class of biologically active substances. *Nature* **183**, 55–56 (1959).
2. Schaal, B. et al. Chemical and behavioural characterization of the rabbit mammary pheromone. *Nature* **424**, 68–72 (2003).
3. Novotny, M. V., Weidong, M., Wiesler, D. & Zidek, L. Positive identification of the puberty-accelerating pheromone of the house mouse: the volatile ligands associating with the major urinary protein. *Proc. R. Soc. Lond. B* **266**, 2017–2022 (1999).
4. Wyatt, T. D. *Pheromones and Animal Behaviour* (Cambridge Univ. Press, Cambridge, 2003).
5. Yamazaki, K., Beauchamp, G. K., Curran, M. & Boyse, E. A. Parent–progeny recognition as a function of MHC odortype identity. *Proc. Natl Acad. Sci. USA* **97**, 10500–10502 (2000).
6. Potts, W. K., Manning, C. J. & Wakeland, E. K. Mating patterns in seminatural populations of mice influenced by MHC genotype. *Nature* **352**, 619–621 (1991).
7. Wysocki, C. & Preti, G. Facts, fallacies, fears and frustrations with human pheromones. *Anat. Rec. A* **281**, 1201–1211 (2004).
8. Lin, D. Y., Zhang, S.-Z., Block, E. & Katz, L. C. Encoding social signals in the mouse main olfactory bulb. *Nature* **434**, 470–477 (2005).
9. Kikuyama, S. et al. Sodefrin: a female-attracting peptide pheromone in newt cloacal glands. *Science* **267**, 1643–1645 (1995).
10. Spehr, M. et al. Essential role of the main olfactory system in social recognition of major histocompatibility complex peptide ligands. *J. Neurosci.* **26**, 1961–1970 (2006).
11. Boehm, T. & Zufall, F. MHC peptides and the sensory evaluation of genotype. *Trends Neurosci.* **29**, 100–107 (2006).
12. Leinders-Zufall, T. et al. MHC class I peptides as chemosensory signals in the vomeronasal organ. *Science* **306**, 1033–1037 (2004).
13. Hurst, J. L. et al. Individual recognition in mice mediated by major urinary proteins. *Nature* **414**, 631–634 (2001).
14. Moré, L. Mouse major urinary proteins trigger ovulation via the vomeronasal organ. *Chem. Senses* **31**, 393–401 (2006).
15. Kimoto, H., Haga, S., Sato, K. & Touhara, K. Sex-specific peptides from exocrine glands stimulate mouse vomeronasal sensory neurons. *Nature* **437**, 898–901 (2005).
16. Sørensen, P. et al. Mixture of new sulfated steroids functions as a migratory pheromone in the sea lamprey. *Nature Chem. Biol.* **1**, 324–328 (2005).
17. Dorries, K. M., Adkins, R. E. & Halpern, B. P. Sensitivity and behavioral responses to the pheromone androstenone are not mediated by the vomeronasal organ in domestic pigs. *Brain Behav. Evol.* **49**, 53–62 (1997).
18. Novotny, M. V. Pheromones, binding proteins and receptor responses in rodents. *Biochem. Soc. Trans.* **31**, 117–122 (2003).
19. Zhang, X. & Firestein, S. The olfactory receptor gene superfamily of the mouse. *Nature Neurosci.* **5**, 124–133 (2002).
20. Liberles, S. D. & Buck, L. B. A second class of chemosensory receptors in the olfactory epithelium. *Nature* **42**, 645–650 (2006).
21. Halpern, M. & Martínez-Marcos, A. Structure and function of the vomeronasal system: an update. *Prog. Neurobiol.* **70**, 245–318 (2003).
22. Dulac, C. & Torello, A. T. Molecular detection of pheromone signals in mammals: from genes to behaviour. *Nature Rev. Neurosci.* **4**, 551–562 (2003).
23. Del Punta, K. et al. Deficient pheromone responses in mice lacking a cluster of vomeronasal receptor genes. *Nature* **419**, 70–74 (2002).
24. Lucas, P., Ukhonov, K., Leinders-Zufall, T. & Zufall, F. A diacylglycerol-gated cation channel in vomeronasal neuron dendrites is impaired in TRPC2 mutant mice: mechanism of pheromone transduction. *Neuron* **40**, 551–561 (2003).
25. Rodríguez, I., Del Punta, K., Rothman, A., Ishii, T. & Mombaerts, P. Multiple new and isolated families within the mouse superfamily of V1r vomeronasal receptors. *Nature Neurosci.* **5**, 134–140 (2002).
26. Yang, H., Shi, P., Zhang, Y. & Zhang, J. Composition and evolution of the V2r vomeronasal receptor gene repertoire in mice and rats. *Genomics* **86**, 306–315 (2005).
27. Hansen, A., Anderson, K. & Finger, T. Differential distribution of olfactory receptor neurons in goldfish: structural and molecular correlates. *J. Comp. Neurol.* **477**, 347–359 (2004).
28. Pfister, R. & Rodríguez, I. Olfactory expression of a single and highly variable V1r pheromone receptor-like gene in fish species. *Proc. Natl Acad. Sci. USA* **102**, 5489–5494 (2005).
29. Eisthen, H. The goldfish knows: olfactory receptor cell morphology predicts receptor gene expression. *J. Comp. Neurol.* **477**, 341–346 (2004).
30. Elsaesser, R., Montani, G., Tirindelli, R. & Paysan, J. Phosphatidylinositol signalling proteins in a novel class of sensory cells in the mammalian olfactory epithelium. *Eur. J. Neurosci.* **21**, 2692–2700 (2005).
31. Meredith, M. Chronic recording of vomeronasal pump activation in awake behaving hamsters. *Physiol. Behav.* **56**, 345–354 (1994).
32. Leinders-Zufall, T. et al. Ultrasensitive pheromone detection by mammalian vomeronasal neurons. *Nature* **405**, 792–796 (2000).
33. Sam, M. et al. Odorants may arouse instinctive behaviours. *Nature* **412**, 142 (2001).
34. Bosch, C. et al. Pheromone detection mediated by a V1r vomeronasal receptor. *Nature Neurosci.* **5**, 1261–1262 (2002).
35. Sharrow, S., Vaughn, J., Zidek, L., Novotny, M. & Stone, M. Pheromone binding by polymorphic mouse major urinary proteins. *Protein Sci.* **11**, 2247–2256 (2002).
36. Luo, M. M., Fee, M. S. & Katz, L. C. Encoding pheromonal signals in the accessory olfactory bulb of behaving mice. *Science* **299**, 1196–1201 (2003).
37. Xu, F. et al. Simultaneous activation of the mouse main and accessory olfactory bulbs by odors or pheromones. *J. Comp. Neurol.* **489**, 491–500 (2005).
38. Wysocki, C. J. & Lepri, J. L. Consequences of removing the vomeronasal organ. *J. Steroid Biochem. Mol. Biol.* **39**, 661–669 (1991).
39. Stowers, L., Holy, T. E., Meister, M., Dulac, C. & Koentges, G. Loss of sex discrimination and male–male aggression in mice deficient for TRP2. *Science* **295**, 1493–1500 (2002).
40. Leybold, B. G. et al. Altered sexual and social behaviors in trp2 mutant mice. *Proc. Natl Acad. Sci. USA* **99**, 6376–6381 (2002).
41. Kelliher, K., Spehr, M., Li, X.-H., Zufall, F. & Leinders-Zufall, T. Pheromonal recognition memory induced by TRPC2-independent vomeronasal sensing. *Eur. J. Neurosci.* **23**, 3385–3390 (2006).
42. Hudson, R. & Distel, H. Pheromonal release of suckling in rabbits does not depend on the vomeronasal organ. *Physiol. Behav.* **37**, 123–129 (1986).
43. Mandiyan, V., Coats, J. & Shah, N. Deficits in aggressive behaviors in *Cnga2* mutant mice. *Nature Neurosci.* **8**, 1660–1662 (2005).
44. Hurst, J. & Beynon, R. Scent wars: the chemobiology of competitive signalling in mice. *BioEssays* **26**, 1288–1298 (2004).
45. Zufall, F. & Munger, S. From odor and pheromone transduction to the organization of the sense of smell. *Trends Neurosci.* **24**, 191–193 (2001).
46. Ma, M. et al. Olfactory signal transduction in the mouse septal organ. *J. Neurosci.* **23**, 317–324 (2003).
47. Fuss, S., Omura, M. & Mombaerts, P. The Grueneberg ganglion of the mouse projects axons to glomeruli in the olfactory bulb. *Eur. J. Neurosci.* **22**, 2649–2654 (2005).
48. Jordan, W. C. & Bruford, M. W. New perspectives on mate choice and the MHC. *Heredity* **81**, 127–133 (1998).
49. Boyse, E. A., Beauchamp, G. K. & Yamazaki, K. The genetics of body scent. *Trends Genet.* **3**, 97–102 (1987).
50. Carroll, L. S., Penn, D. J. & Potts, W. K. Discrimination of MHC-derived odors by untrained mice is consistent with divergence in peptide-binding region residues. *Proc. Natl Acad. Sci. USA* **99**, 2187–2192 (2002).
51. Singh, P. B., Brown, R. E. & Roser, B. MHC antigens in urine as olfactory recognition cues. *Nature* **327**, 161–164 (1987).
52. Singh, P. B. Chemosensation and genetic individuality. *Reproduction* **121**, 529–539 (2001).
53. Schaefer, M. L., Yamazaki, K., Osada, K., Restrepo, D. & Beauchamp, G. K. Olfactory fingerprints for major histocompatibility complex-determined body odors II: relationship among odor maps, genetics, odor composition, and behavior. *J. Neurosci.* **22**, 9513–9521 (2002).
54. Loconto, J. et al. Functional expression of murine V2R pheromone receptors involves selective association with the M10 and M1 families of MHC class Ib molecules. *Cell* **112**, 607–618 (2003).
55. Ishii, T., Hirota, J. & Mombaerts, P. Combinational coexpression of neural and immune multigene families in mouse vomeronasal sensory systems. *Curr. Biol.* **13**, 394–400 (2003).
56. Olson, R., Huey-Tubman, K., Dulac, C. & Bjorkman, P. Structure of a pheromone receptor-associated MHC molecule with an open and empty groove. *PLoS* **3**, e257 (2005).
57. Milinski, M. et al. Mate choice decisions of stickleback females predictably modified by MHC peptide ligands. *Proc. Natl Acad. Sci. USA* **102**, 4414–4418 (2005).
58. Flower, D. R. The lipocalin protein family: structure and function. *Biochem. J.* **318**, 1–14 (1996).
59. Armstrong, S., Robertson, D., Cheetham, S., Hurst, J. & Beynon, R. Structural and functional differences in isoforms of mouse major urinary proteins: a male-specific protein that preferentially binds a male pheromone. *Biochem. J.* **391**, 343–350 (2005).
60. Beynon, R. et al. Polymorphism in major urinary proteins: molecular heterogeneity in a wild mouse population. *J. Chem. Ecol.* **28**, 1429–1446 (2002).
61. Luo, M. & Katz, L. Encoding pheromonal signals in the mammalian vomeronasal system. *Curr. Opin. Neurobiol.* **14**, 428–434 (2004).
62. Sugai, T., Sugitani, M. & Onoda, N. Subdivisions of the guinea pig accessory olfactory bulb revealed by the combined method with immunohistochemistry, electrophysiological, and optical recordings. *Neuroscience* **79**, 871–885 (1997).
63. Boehm, U., Zou, Z. & Buck, L. Feedback loops link odor and pheromone signaling with reproduction. *Cell* **123**, 683–695 (2005).
64. Yoon, H., Enquist, L. & Dulac, C. Olfactory inputs to hypothalamic neurons controlling reproduction and fertility. *Cell* **123**, 669–695 (2005).
65. Kaba, H. & Nakanishi, S. Synaptic mechanisms of olfactory recognition memory. *Rev. Neurosci.* **6**, 125–141 (1995).
66. Halem, H. A., Cherry, J. A. & Baum, M. J. Central forebrain responses to familiar male odors are attenuated in recently mated female mice. *Eur. J. Neurosci.* **13**, 389–399 (2001).
67. Binns, K. E. & Brennan, P. A. Changes in electrophysiological activity in the accessory olfactory bulb and medial amygdala associated with mate recognition in mice. *Eur. J. Neurosci.* **21**, 2529–2537 (2005).
68. Yamazaki, K. et al. Familial imprinting determines H-2 selective mating preferences. *Science* **240**, 1331–1332 (1988).
69. Penn, D. J. & Potts, W. MHC-disassortative mating preferences reversed by cross-fostering. *Proc. R. Soc. Lond. B* **265**, 1299–1306 (1998).
70. Hudson, R. Do newborn rabbits learn the odor stimuli releasing nipple-search behavior? *Dev. Psychobiol.* **18**, 575–585 (1985).
71. Moncho-Bogani, J., Martínez-García, F., Novejarque, A. & Lanuza, E. Attraction to sexual pheromones and associated odors in female mice involves activation of the reward system and basolateral amygdala. *Eur. J. Neurosci.* **21**, 2186–2198 (2005).
72. Meredith, M. Vomeronasal, olfactory, hormonal convergence in the brain. *Ann. NY Acad. Sci.* **855**, 349–361 (1998).
73. Mitchell, J. & Gratton, A. Opioid modulation and sensitization of dopamine release elicited by sexually relevant stimuli: a high speed chronoamperometric study in freely behaving rats. *Brain Res.* **551**, 20–27 (1991).
74. West, C., Clancy, A. & Michael, R. Enhanced responses of nucleus accumbens neurons in male rats to novel odors associated with sexually receptive females. *Brain Res.* **585**, 49–55 (1992).
75. Kippin, T., Cain, S. & Pfau, J. Estrous odors and sexually conditioned neutral odors activate separate neural pathways in the male rat. *Neuroscience* **117**, 971–979 (2003).
76. Meredith, M. & Westberry, J. M. Distinctive responses in the medial amygdala to same-species and different-species pheromones. *J. Neurosci.* **24**, 5719–5725 (2004).
77. Choi, G. B. et al. Lhx6 delineates a pathway mediating innate reproductive behaviors from the amygdala to the hypothalamus. *Neuron* **46**, 647–660 (2005).
78. Ferguson, J. N., Aldag, J. M., Insel, T. R. & Young, L. J. Oxytocin in the medial amygdala is essential for social recognition in the mouse. *J. Neurosci.* **21**, 8278–8285 (2001).
79. Keller, M., Perrin, G., Meurisse, M., Ferreira, G. & Lévy, F. Cortical and medial amygdala are both involved in the formation of olfactory offspring memory in sheep. *Eur. J. Neurosci.* **20**, 3433–3441 (2004).
80. Leyden, J., McGinley, K., Hölzle, E., Labows, J. & Kligman, A. The microbiology of the human axilla and its relationship to axillary odor. *J. Invest. Dermatol.* **77**, 413–416 (1981).
81. Zeng, X.-N. et al. Analysis of characteristic odors from human male axillae. *J. Chem. Ecol.* **17**, 1469–1492 (1991).

82. Zeng, C. *et al.* A human axillary odorant is carried by apolipoprotein D. *Proc. Natl Acad. Sci. USA* **93**, 6626–6630 (1996).
83. Witt, M. & Hummel, T. Vomeronasal versus olfactory epithelium: is there a cellular basis for human vomeronasal perception? *Int. Rev. Cytol.* **248**, 209–259 (2006).
84. Meredith, M. Human vomeronasal organ function: a critical review of best and worst cases. *Chem. Senses* **26**, 433–445 (2001).
85. Zhang, J. & Webb, D. Evolutionary deterioration of the vomeronasal pheromone transduction pathway in catarrhine primates. *Proc. Natl Acad. Sci. USA* **100**, 8337–8341 (2003).
86. Liman, E. R. & Innan, H. Relaxed selective pressure on an essential component of pheromone transduction in primate evolution. *Proc. Natl Acad. Sci. USA* **100**, 3328–3332 (2003).
87. Kouros-Mehr, H. *et al.* Identification of non-functional human VNO receptor genes provides evidence for vestigiality of the human VNO. *Chem. Senses* **26**, 1167–1174 (2001).
88. Rodriguez, I. & Mombaerts, P. Novel human vomeronasal receptor-like genes reveal species-specific families. *Curr. Biol.* **12**, R409–R411 (2002).
89. Rodriguez, I., Greer, C. A., Mok, M. Y. & Mombaerts, P. A putative pheromone receptor gene expressed in human olfactory mucosa. *Nature Genet.* **26**, 18–19 (2000).
90. Stern, K. & McClintock, M. K. Regulation of ovulation by human pheromones. *Nature* **392**, 177–179 (1998).
91. Preti, G., Wysocki, C., Barnhart, K., Sondheimer, S. & Leyden, J. Male axillary extracts contain pheromones that affect pulsatile secretion of luteinizing hormone and mood in women recipients. *Biol. Reprod.* **68**, 2107–2113 (2003).
92. Wedekind, C. & Furi, S. Body odour preferences in men and women: do they aim for specific MHC combinations or simply heterozygosity? *Proc. R. Soc. Lond. B* **264**, 1471–1479 (1997).
93. Jacob, S., McClintock, M. K., Zelano, B. & Ober, C. Paternally inherited HLA alleles are associated with women's choice of male odor. *Nature Genet.* **30**, 175–179 (2002).
94. Ober, C. *et al.* HLA and mate choice in humans. *Am. J. Hum. Genet.* **61**, 497–504 (1997).
95. Varendi, H. & Porter, R. H. Breast odour as the only maternal stimulus elicits crawling towards the odour source. *Acta Paediatr.* **90**, 372–375 (2001).
96. Jacob, S. & McClintock, M. Psychological state and mood effects of steroidal chemosignals in women and men. *Horm. Behav.* **37**, 57–78 (2000).
97. Savic, I., Berglund, H., Gulyas, B. & Roland, P. Smelling of odorous sex hormone-like compounds causes sex-differentiated hypothalamic activations in humans. *Neuron* **31**, 661–668 (2001).
98. Keverne, E. Odor here, odor there: chemosensation and reproductive function. *Nature Neurosci.* **8**, 1637–1638 (2005).
99. Brennan, P. A., Kendrick, K. M. & Keverne, E. B. Neurotransmitter release in the accessory olfactory bulb during and after the formation of an olfactory memory in mice. *Neuroscience* **69**, 1075–1086 (1995).
100. Pitkänen, A. in *The Amygdala: a Functional Analysis* (ed. Aggleton, J.) 31–115 (Oxford University Press, Oxford, 2000).

Acknowledgements In memory of L. C. Katz. This review and our research were supported by grants from the NIH/NIDCD (F.Z.) and the MRC/BBSRC (P.A.B.). Apologies to our colleagues whose work we could not cite owing to space limitations.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence should be addressed to the authors (p.brennan@bristol.ac.uk; fzufa001@umaryland.edu).

Smell images and the flavour system in the human brain

Gordon M. Shepherd¹

Flavour perception is one of the most complex of human behaviours. It involves almost all of the senses, particularly the sense of smell, which is involved through odour images generated in the olfactory pathway. In the human brain, the perceptual systems are closely linked to systems for learning, memory, emotion and language, so distributed neural mechanisms contribute to food preference and food cravings. Greater recognition of the role of the brain's flavour system and its connection with eating behaviour is needed for a deeper understanding of why people eat what they do, and to generate better recommendations about diet and nutrition.

Many humans regard their sense of smell to be of minor importance. The dominant role that smell has in sensing the flavours of the foods we eat and influencing what we like to eat is largely unrecognized. This is unfortunate, because our diet is causing a public health crisis in many western countries. Knowledge of the importance of smell to the perception of flavour and the formation of cognitive and emotional responses to foods and beverages could help to improve health and prevent chronic conditions such as obesity and diabetes.

The aim of this article is to review recent advances in the brain mechanisms of smell perception and to put forward several hypotheses for integrating these mechanisms into current theories of the neural basis of flavour as a complex multisystem percept. My hope is that this will help bridge the divide that impedes neuroscientists, behavioural psychologists, food scientists, nutritionists and public-policy makers from communicating effectively across disciplines about why we eat what we do and how to achieve healthy diets. I focus on three critical aspects of the brain mechanisms of smell: how odour molecules are represented in the form of 'odour images' as the basis for smell perception; the importance of retronasal smell for flavour perception; and the close relationships between the brain mechanisms for flavour and the brain systems for emotions and cravings.

Odour molecules are represented by odour images

Odour stimulation of olfactory sensory cells in the nose involves odour molecules interacting with olfactory-receptor proteins. The discovery of the genes that encode these proteins in 1991 by Richard Axel and Linda Buck¹ moved olfaction into the front rank among scientific endeavours, not only opening the way to understanding olfactory (and taste) transduction at the receptor level, but also providing new genetic and molecular tools for pursuing the organization of the olfactory pathway in the brain.

The genome projects show that the family of olfactory genes totals more than 1,000 — the largest in the mammalian genome². In the majority of mammals, most of the genes are functional, but in primates the number of functional genes decreases, in humans to about 350 (archived in the Olfactory Receptor Database; see ref. 3). This has been taken to confirm the belief that humans have degenerate senses of smell. However, olfactory genes do not map directly onto smell acuity; dogs, which have superior tracking abilities, have only about 850 functional genes⁴.

In fact, behavioural tests show that primates have surprisingly good senses of smell⁵, and it has been argued that the decline in olfactory-gene number is more than offset in humans by their much enlarged brains with their increased capacities for analysis and complex processing of smell to guide critical human behaviours⁶. As we shall see, a notable share of this enlarged brain capacity is involved in flavour perception and behavioural responses to flavours.

The axons of sensory neurons in the olfactory epithelium project to the olfactory bulb, where they converge onto modules known as glomeruli (Fig. 1a). Direct evidence of the brain mechanisms involved in smell first emerged in the 1970s with the application of the earliest brain-imaging methods to the olfactory bulb. This showed that odour stimuli in the rat produce spatial patterns of activity in the glomerular layer that are overlapping but that vary for different odours⁷. From this was derived the fundamental principle, first suggested by Adrian⁸, that different odour molecules are represented by different patterns of spatial activity in the olfactory bulb. A homologous chemical series — such as aldehydes — elicits patterns that are overlapping but different, as shown by 2-deoxyglucose⁹ and high-resolution functional magnetic resonance imaging (fMRI)¹⁰ (Fig. 1b). Behavioural studies¹¹ have confirmed that mice can readily discriminate such single-carbon differences between molecules.

The olfactory system thus resembles other sensory pathways in that it uses activity patterns in two-dimensional neural space to represent its sensory modality. And such patterns, analogous to those of other sensory systems, constitute 'odour images', or 'odour maps'. Much research has confirmed and extended this basic finding in the olfactory system (for a review see ref. 12). Some recent studies have focused on characterizing how these images evolve during a respiratory cycle¹³, whereas others have analysed the contribution of temporal firing patterns to the encoding process¹⁴. Work on tracing the connections in the olfactory bulb is revealing the distributed circuits that process the images¹⁵. These allow two-dimensional images (Fig. 1b) to encode the multidimensional odour space inherent in the complex structures of odour molecules. This principle of odour-molecule representation by glomerular activity patterns applies across phyla¹⁶.

Despite this emerging consensus, the idea of odour images and odour maps has not yet been incorporated into current theories on the neural basis of smell and flavour perception (but see ref. 17), or in psychophysical

¹Department of Neurobiology, Yale University School of Medicine, 333 Cedar Street, New Haven, Connecticut 06510, USA.

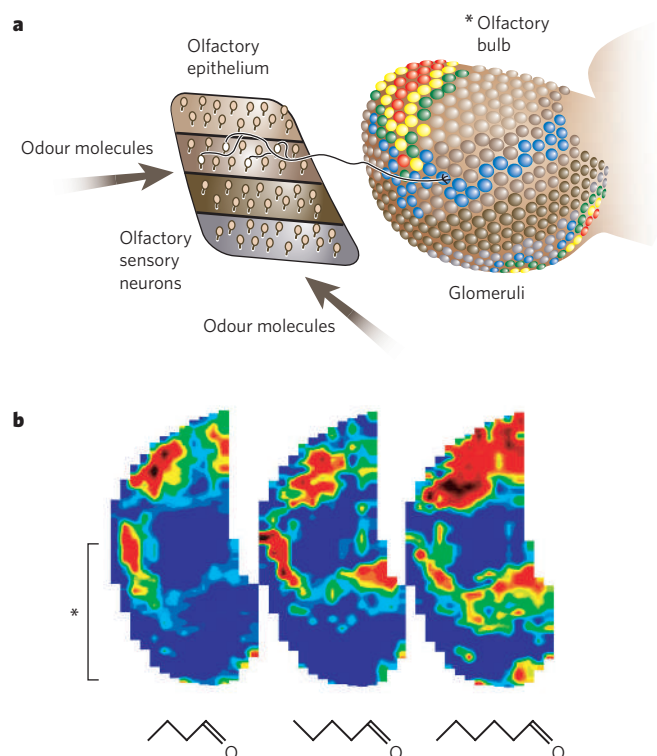


Figure 1 | Odour images in the olfactory glomerular layer. **a**, Diagram showing the relationship between the olfactory receptor cell sheet in the nose and the glomeruli of the olfactory bulb⁵³. **b**, fMRI images of the different but overlapping activity patterns seen in the glomerular layer of the olfactory bulb of a mouse exposed to members of the straight-chain aldehyde series, varying from four to six carbon atoms. The lower part of the image in the left panel corresponds to the image on the medial side of the olfactory glomerular layer as shown in **a** (see asterisk). (Image in **a** adapted, with permission, from ref. 53; image in **b** adapted, with permission, from ref. 10.)

studies of the relationship between odour molecule structure and perceptual response. This resistance seems to be due to at least two factors.

First, unlike the conscious images we perceive through the visual system, the odour patterns that arise in the olfactory bulb are unconscious. When we perceive a smell we do not think of it as a spatial image, even though it arose from a spatial image in the olfactory bulb. This unconscious nature may reflect the fact that the human olfactory pathway does not include a significant relay in the thalamus (see below).

Second, odour patterns are complex and highly irregular (Fig. 1). Recognition of odour images is thus an example of pattern recognition. We have suggested that this is analogous to recognizing complex visual patterns such as those of human faces⁶. Just as the identity of a face is encoded in its visual image, so is the identity of an odour molecule or complex odour object — such as the scent of a flower or the aroma of coffee — encoded in an odour image.

Unconscious pattern recognition of odour images therefore seems to lie at the heart of smell perception. This view is being incorporated by food scientists into their concepts of the neural basis of flavour¹⁸. Neuroscientists and psychophysicists have the opportunity to follow suit and devise experiments that test this hypothesis, which we will use to describe the role of smell in flavour perception.

Retronasal smell is essential for human flavour perception

As noted by Paul Rozin¹⁹, smell is unique in having a ‘dual nature’ — that is, it can sense signals originating outside (orthonasal) and inside (retronasal) the body (Table 1).

Orthonasal stimulation refers to sniffing in through the external nares of the nose to activate the sensory cells in the olfactory epithelium (Fig. 2a).

This is the route used to sense odours in the environment (Table 1). The resultant odour images in the olfactory bulb are subject to processing by the olfactory cortex and are relayed to the primary olfactory cortex in the orbitofrontal cortex (OFC), a part of the prefrontal lobe. This direct input to the highest cognitive centres of the brain is a special property of smell that is critical to the experience of human flavour.

Retronasal stimulation occurs during food ingestion, when volatile molecules released from the food in the mouth are pumped, by movements of the mouth, from the back of the oral cavity up through the nasopharynx to the olfactory epithelium. Orthonasal olfaction does not contribute to this sensation. Retronasal olfactory stimulation can be studied by introducing odour stimuli at the back of the mouth²⁰ or by their release from ingested foods²¹. Marcel Proust vividly described how a tea-soaked madeleine brought back powerful childhood memories²², and this would have occurred primarily through the retronasal pathway. It is activated only when breathing out through the nose between mastications or swallowings²³ (Fig. 2b). Because the molecules arise from the food in the mouth, they are referred to the mouth; that is, they are perceived as if they are sensed within the mouth²⁴. This retronasal food-molecule-laden air is judged to be part of the ‘taste’ of the food²⁵ and has been shown to be necessary for flavour identification²⁶. Thus, although a large part of flavour is due to smell, it is attributed to ‘taste’, and is therefore almost entirely hidden from our conscious perception. To avoid giving ‘taste’ a double meaning, we incorporate retronasal smell as one of the modalities covered by the term ‘flavour’.

Retronasal smell and the brain flavour system

The traditional view in the literature on eating behaviour in human culture²⁷ is that the flavour of prepared foods is humanity’s greatest universal shared behaviour, experienced by individuals of all ages in the course of daily life. Flavour is also among the most complex and powerful of all human sensations. It engages almost all of the sensory modalities. It also engages the complex facial, swallowing and respiratory motor systems. Flavour is therefore an active sensation — we use ‘active taste’ to palpate our food with our tongue as we use ‘active touch’ to palpate an object we are examining with our fingers. Some of these systems are indicated in the diagram of Fig. 2b. Above these sensory and motor systems are the cognitive systems for memory, emotion, abstract thinking and language. The importance of retronasal smell images is illustrated by the massive extent to which they interact with these brain systems compared with orthonasal smell images (compare Fig. 2a with Fig. 2b). Below, we briefly consider the main sensory systems.

Taste is the most obvious component of food flavour. The receptors that relate to sweet, salt, sour, bitter and umami taste are described in detail by Zuker in this issue (page 288). Here, I touch only on certain aspects of taste relevant to the role of smell in flavour. In humans, the taste pathway ascends from the nucleus of the solitary tract in the brainstem to the hypothalamus and to the taste area of the somatosensory thalamus, and

Table 1 | The dual olfactory system

Operations	Orthonasal olfaction	Retronasal olfaction
Stimulation route	Through the external nares	From the back of the mouth through the nasopharynx
Stimuli	Floral scents Perfumes Smoke Food aromas Prey/predator smells Social odors Pheromones MHC molecules	Food volatiles
Processed by	Olfactory pathway influenced by the visual pathway	Olfactory pathway combined with pathways for taste, touch, sound and active sensing by proprioception form a ‘flavour system’

Note the interesting contrast, that orthonasal olfactory perception involves a wide range of types of odors processed through only the olfactory pathway, in comparison with retronasal olfactory perception which involves only food volatiles but processed in combination with many brain pathways.

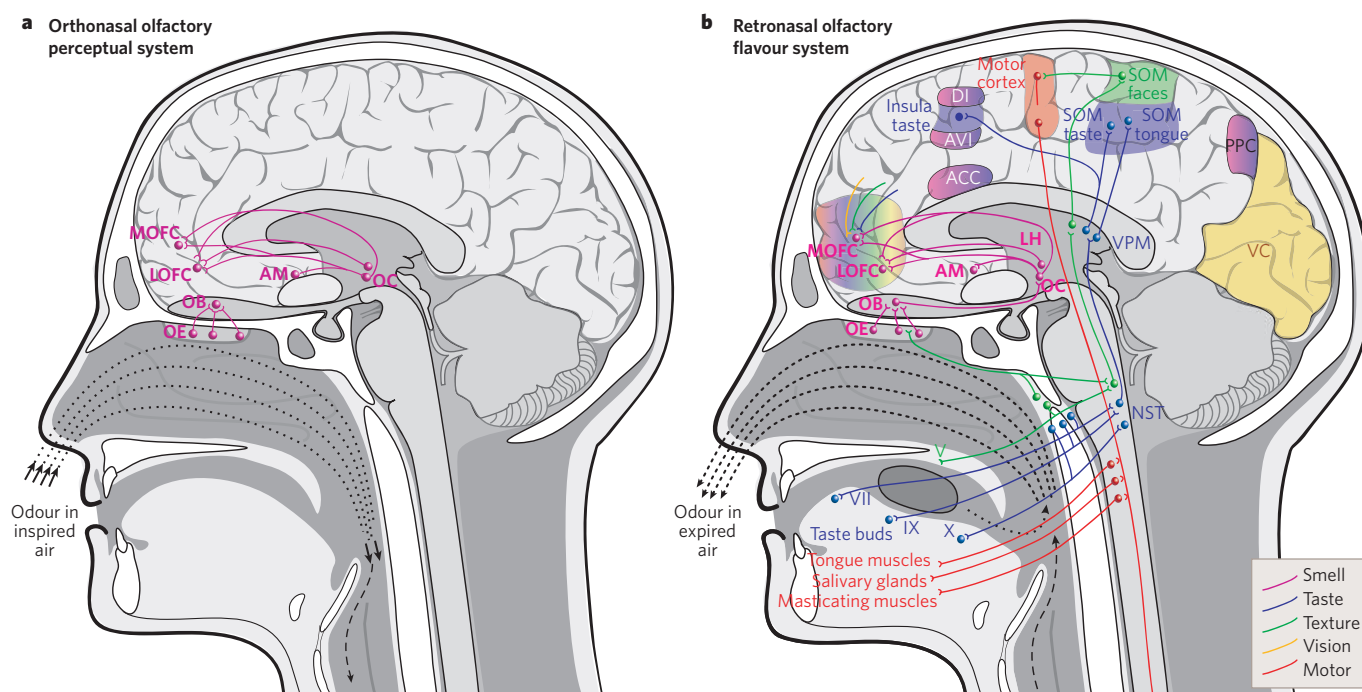


Figure 2 | The dual olfactory system. **a**, Brain systems involved in smell perception during orthonasal olfaction (sniffing in). **b**, Brain systems involved in smell perception during retronasal olfaction (breathing out), with food in the oral cavity. Air flows indicated by dashed and dotted lines; dotted lines indicate air carrying odour molecules. ACC, accumbens; AM, amygdala; AVI, anterior ventral insular cortex; DI, dorsal insular cortex; LH, lateral hypothalamus; LOFC, lateral orbitofrontal cortex; MOFC, medial orbitofrontal cortex; NST, nucleus of the solitary tract; OB, olfactory bulb; OC, olfactory cortex; OE, olfactory epithelium; PPC, posterior parietal cortex; SOM, somatosensory cortex; V, VII, IX, X, cranial nerves; VC, primary visual cortex; VPM, ventral posteromedial thalamic nucleus.

from there extends to the primary taste cortex (Fig. 2b). A key fact about taste stimuli is that they elicit the most basic human emotions of pleasure (sweet) and disgust (bitter), which are not learned; they are hard-wired in the brainstem from birth^{28,29}. In contrast to taste, the affective responses to odour images seem to be mostly learned, which presumably accounts for the enormous diversity of flavours in the world's cuisines.

Using brain imaging, researchers have studied how smell and taste pathways function, separately and together, in humans³⁰. At the neocortical level, taste stimuli alone activated the primary taste centres in the tongue area and insula, whereas odour stimuli by themselves activated the primary olfactory receiving area in the orbitofrontal region of prefrontal cortex (Fig. 2b), as expected. The investigators then tested for simultaneous presentation of the taste and odour stimuli. This produced enhanced activity in many of the regions activated by the independent stimulations as well as additional activity in contiguous areas around the primary receiving areas (Fig. 2b). This indicates that a wider range of integrative activity underlies the more complex perception of flavour than would occur through simple addition of the taste and smell pathways.

By contrast, neocortical integration in rodents is minimal. The important principle is that in humans the increase in processing at this higher cognitive level is independent of the smaller size of the receptor repertoire. A possible analogy is with language: the development of human language was not dependent on more hair cells or a more acute sense of hearing, but on the elaboration of neocortical areas for the complex processing of low- to medium-frequency auditory signals. Similarly, it can be proposed that the increased processing and elaboration of flavour through these larger cognitive regions is unique to humans (Box 1). Processing in this case encompasses odour images together with the other senses that contribute to flavour. But how does this occur?

The orbitofrontal sensory packaging centre

The prefrontal cortex has long been recognized as a special cortical area in the human brain that is crucial for planning, abstract thought and

working memory. As indicated in Fig. 2b, it is also particularly important in flavour perception. Anatomical studies³¹ in monkeys show that in addition to the primary neocortical olfactory area there are various subregions differentiated to receive connections from other sensory neocortical areas, including taste, hearing, touch and vision. There are also

Box 1 | Smell images, flavours and language

It is commonly assumed that there is a remote relationship between language and smell; for example, it is difficult to find words to describe the taste of a wine. It has been postulated that odours are difficult to name because odour and language processing rely on the same neural substrates⁵⁶. Olfactory perceptions are also heavily conditioned by verbal cues, in the same way that they are influenced by other senses⁵⁷. However, the naming of smells recognized through odour images may be no more difficult than naming of visual images such as faces; both involve the inherent difficulty of describing complex spatial patterns in words. We are very good at recognizing human faces but have great difficulty describing them, and it is similarly difficult for us to describe a smell.

The relationship between flavour and language is even more interesting when retronasal odour images are considered to be a main component of food flavours. Cooking had a central role in human evolution⁵⁸; the shared common meal has long been regarded as the defining social activity of early humans²⁷. Within this social context, language arose not only in relation to the activities of hunting and gathering, but also as a necessary means to communicate about the preparation of the meal and its assessment in terms of desirable flavours. It can be proposed that the adaptive advantage of language allowed humans to attach verbal labels to the odour images associated with new flavours produced by cooking, facilitating the process of extracting more nutritious food from otherwise unpalatable foodstuffs such as roots, nuts or toxic plants⁵⁸, or for the cultural advantage it conferred in the competition between rival human groups.

subregions with reciprocal connections with limbic subcortical regions such as the amygdala and the hypothalamus. An extensive series of physiological studies³² in monkeys with recordings from single cells in these areas has shown that responses to taste and both orthonasal and retro-nasal stimuli are heavily dependent on context, especially in connection with learning and reward. Similar results have been shown by fMRI studies in humans^{33,34}. These suggest that taste and, in particular, smell perceptions are more heavily conditioned by multi-sensory inputs than by perceptions in other systems such as those for vision and hearing.

Integral components of our experience of eating also arise from all submodalities of the somatosensory system: fine touch, creaminess, deep pressure (for example, crunchy), temperature and pain (such as the burn of chillies). Astringency, such as that of a dry wine, also gives a sense of body.

These systems and their submodalities are represented in Fig. 3, and their pathways from the sensors to the receiving cortical regions are indicated in Fig. 2b. Flavour is also influenced by the sight of the food to be consumed³⁵. A recent study³⁶ tested both naive and experienced subjects on red and white wines. Both groups classified the wines according to the contrasting tastes of red and white wines. However, the reds included whites that had been coloured red, showing that the colour had an overriding effect on the perceived flavour. The visual cortex is therefore included in Fig. 2b as a part of the greater flavour system (see also Table 1).

In summary, the multisystem nature of the flavour centre of the OFC (Fig. 2b) suggests that this can be regarded as a sort of 'sensory packaging centre' in the brain, one of the areas that makes us susceptible to the pitch of advertisers to buy food in different coloured boxes or with more crunchy texture. This centre has a critical and asymmetric relationship

with human language areas (Fig. 3). On the one hand, there is the difficulty humans have describing in words the spatiotemporal patterns of odour images as part of flavour perception. On the other hand, there is the strong influence of words on our flavour perceptions and preferences (Box 1).

From images of smell to images of desire

The perception of flavour thus involves many sensory and motor systems. But in order for a flavour to play a part in food preference and consumption, it must also connect with other systems that are involved in the complex control of feeding. Recent research has begun to show how flavour perception might drive these systems as well.

The olfactory cortex and its interactions with other limbic areas form a central hub in this process (Figs 2b, 3). In the olfactory cortex, the odour images are reformatted into content-addressable memory form^{37–40}. (The olfactory cortex also contains a sensor for essential amino acids⁴¹, which is indicated by the asterisk in Fig. 3.) A number of studies have identified the sub-areas that are activated by food flavours. These include the OFC, parahippocampal gyrus, anterior fusiform gyrus, insula, striatum and cingulate. These are mostly cortical areas, or centres closely associated with them. Moreover, the OFC and amygdala have been found to have associations with the desirability of foods, with pleasurable foods activating the medial OFC and unpleasant foods activating the lateral OFC⁴². It is therefore possible to begin to identify the different feelings and states of motivation about feeding in relation to levels of activation of the extended amygdalar system, from a level of liking or disliking a flavour to feeling strongly attracted or repelled by it (Fig. 3).

The motivation to eat is also involved in cravings, overeating and conditions such as obesity, in which energy intake chronically exceeds

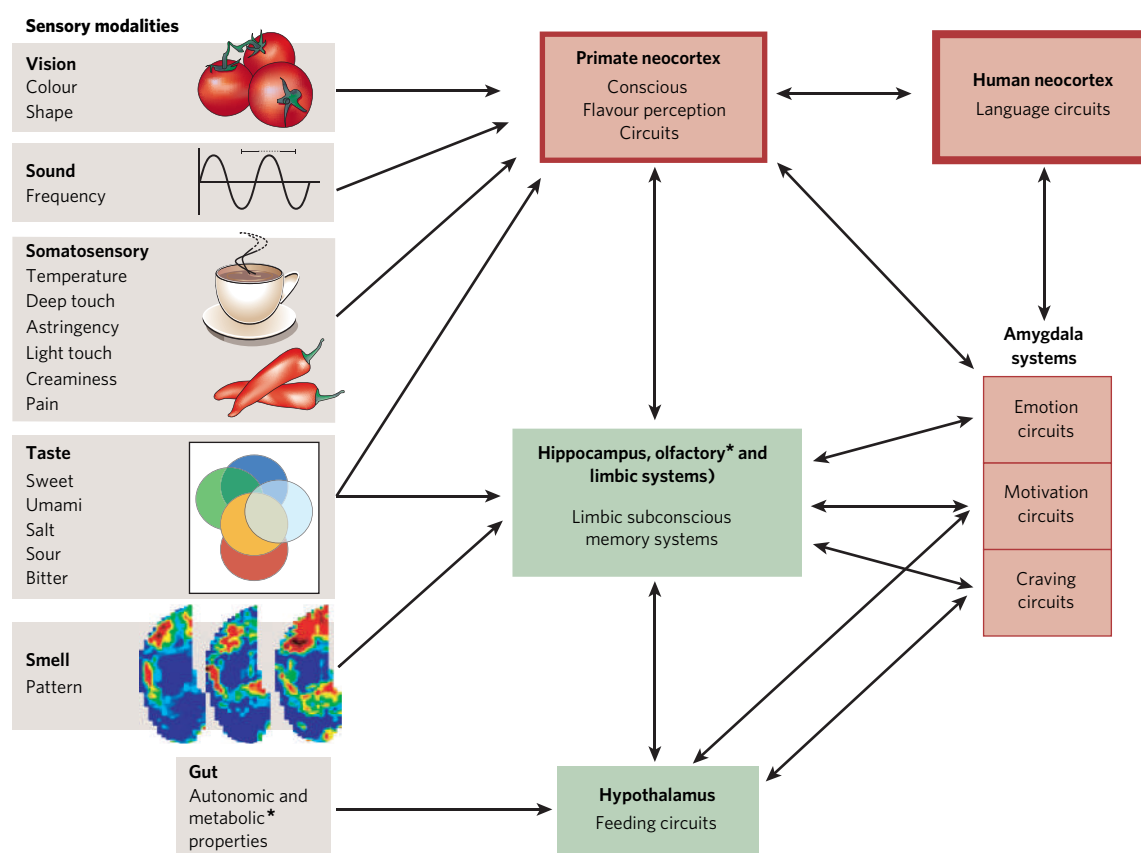


Figure 3 | The human brain flavour systems that evaluate and regulate food intake. The diagram shows the areas involved in the perceptual, emotional, memory-related, motivational and linguistic aspects of food evaluation mediated by flavour inputs^{32,41,54,55}. Left, different sensory modalities and submodalities that contribute to flavour perception. Middle and right, brain flavour system that evaluates and regulates food intake. Red regions mediate conscious sensory perception; thicker outlines indicate their greater importance in humans and other primates. Green regions mediate subconscious feeding regulation. Deficiencies in essential amino acids are sensed by the anterior olfactory cortex (asterisk).

energy expenditure. Various animal studies have shown that the regions involved in cravings include the amygdala, anterior cingulate, OFC, insula, hippocampus, caudate and dorsolateral prefrontal cortex. Functional imaging studies in humans have tested for activity in these areas⁴³ by maintaining subjects on a monotonous diet and asking them to imagine the taste, smell and texture (in other words, flavour) of their favourite foods. Craving-related increases in the fMRI blood-oxygen-level dependent (BOLD) signal were seen in the hippocampus, insula and caudate nucleus. The authors characterized these activity sites as 'images of desire'.

The regions activated in humans by food cravings are similar to those activated during drug and alcohol addiction. The results are consistent with the common substrates hypothesis — the idea, arising from many converging lines of work, that the brain areas underlying food cravings share cellular and systems mechanisms with those involved in alcohol and drug abuse^{44,45}.

Thus, the images of smells in the olfactory pathway lead to images of desire for acquiring and ingesting the object supplying that smell image, and so, it can be proposed, the cycle repeats itself. The key point is that understanding overeating and obesity involves not only understanding the hypothalamic feeding centres and how they respond to fats, carbohydrates and proteins (Fig. 3), but how those centres are driven by the brain mechanisms underlying the flavours of those foods and the desire to consume them.

Food science and beyond

Behavioural testing of the effects of food flavours and consumer preference for them is carried out in food science, a field almost entirely unknown to neuroscientists on the one hand and nutritionists on the other. Using combinations of gas chromatographic methods linked to odour testing, these studies are providing essential links between the physical and chemical properties of foods and the flavour perceptions they produce. The level of detail can be exquisite; one such study, for example, analysed the type of grass consumed by cows hourly during the day and in different seasons, and how this related to the chemical composition and flavour qualities of cheese produced from the cows' milk⁴⁶. Another study investigated the change in retronasal volatiles released during successive stages of ingestion, mastication and swallowing⁴⁷. Such research is contributing to the effort to enhance traditional methods of producing foods with naturally rich flavours⁴⁸, as well as to the effort to maintain or simulate these qualities in produce for mass markets. Recent studies have also assessed the decline that occurs in flavour perception in humans as they age, when retronasal olfaction seems more vulnerable to loss than orthonasal olfaction⁴⁹.

Human flavour systems in the brain are also of interest in relation to the popular movement, which began in the 1980s, to put the cooking of food on a scientific basis. Explaining the chemistry and physics of food preparation, and the biochemical basis of food flavours, is a lively area of popular communication of food science that is associated with Shirley Corriher⁵⁰ and Harold McGee⁵¹, and termed 'molecular gastronomy' by others⁵². It is possible to imagine a future 'neurogastronomy', combining the biochemistry of food preparation, the molecular biology of the olfactory receptors, and the knowledge of odour images and the brain flavour system reviewed here. Such a synthesis could have the potential to improve our understanding of the human experience of eating and ways to shape it towards more flavourful and healthier diets.

Challenges

The increasing evidence that odour molecules are encoded as odour images provides testable hypotheses for further research. Building on current animal studies, applications of genetic and molecular tools to the large receptor gene family will provide the molecular basis of the organization of the olfactory pathway from the sensory cells to the olfactory cortex, and perhaps, ultimately, to the neocortical level as well. This will merge with progress in the physiological and brain imaging studies mentioned here to give a comprehensive understanding of the neural basis of odour images during orthonasal olfaction.

Extending these studies to retronasal olfaction and the basis of flavour will require more concerted efforts by neuroscientists to integrate olfaction with taste and the other sensory systems, and with psychophysical and ingestive behaviour studies. In this multidisciplinary approach, the expertise of food scientists and nutritionists will also be needed. In addition, anthropologists need to provide a better understanding of the role of smell in the development of cuisines during human evolution.

This integrated approach should make it possible for researchers to attack the difficult questions about human flavour perception. What is the relationship between flavour and nutrition? This would have been critical for human evolution, and is at the core of the current concern about healthy diets. Why do fast foods and certain flavours attract humans to the point of craving? If the smell of the flavour is an image, the work reviewed here suggests that the smell image plays a large part in that attraction. Is a fast-food flavour an attractive 'caricature' of the odour image of the natural food, in the same way that Mickey Mouse is an attractive caricature of a real mouse?

The traditional view is that we are attracted to foods by their biochemical composition, especially by those constituents that make them sweet or fat; recently, attention has shifted to the biochemical actions of food constituents on addiction centres in the brain. Here we have reviewed evidence that flavour is a major factor. If we recognize the power that conscious images have on our daily lives, the power of the unconscious odour images that dominate flavours seems worth investigating. ■

1. Buck, L. & Axel, R. A novel multigene family may encode odorant receptors: a molecular basis for odor recognition. *Cell* **65**, 175–187 (1991).
2. Zhang, X. & Firestein, S. The olfactory receptor gene superfamily of the mouse. *Nature Neurosci.* **5**, 124–133 (2002).
3. Crasto, C., Marenco, L., Miller, P. & Shepherd, G. Olfactory receptor database: a metadata-driven automated population from sources of gene and protein sequences. *Nucleic Acids Res.* **30**, 354–360 (2002).
4. Quignon, P. *et al.* The dog and rat olfactory receptor repertoires. *Genome Biol.* **6**, R83 (2005).
5. Laska, M., Seibt, A. & Weber, A. 'Microsmatic' primates revisited: olfactory sensitivity in the squirrel monkey. *Chem. Senses* **25**, 47–53 (2000).
6. Shepherd, G. M. The human sense of smell: are we better than we think? *PLoS Biol.* **2**, e146 (2004).
7. Stewart, W. B., Kauer, J. S. & Shepherd, G. M. Functional organization of rat olfactory bulb analysed by the 2-deoxyglucose method. *J. Comp. Neurol.* **185**, 715–734 (1979).
8. Adrian, E. D. Sensory messages and sensation: the response of the olfactory organ to different smells. *Acta Physiol. Scand.* **29**, 5–14 (1953).
9. Johnson, B. A., Farahbod, H., Saber, S. & Leon, M. Effects of functional group position on spatial representations of aliphatic odorants in the rat olfactory bulb. *J. Comp. Neurol.* **483**, 192–204 (2005).
10. Xu, F. Q. *et al.* Odor maps of aldehydes and esters revealed by fMRI in the glomerular layer of the mouse olfactory bulb. *Proc. Natl Acad. Sci. USA* **100**, 11029–11034 (2003).
11. Laska, M., Joshi, D. & Shepherd, G. M. Olfactory sensitivity for aliphatic aldehydes in CD-1 mice. *Behav. Brain Res.* **167**, 349–354 (2006).
12. Xu, F., Greer, C. A. & Shepherd, G. M. Odor maps in the olfactory bulb. *J. Comp. Neurol.* **422**, 489–495 (2000).
13. Spors, H., Wachowiak, M., Cohen, L. B. & Friedrich, R. W. Temporal dynamics and latency patterns of receptor neuron input to the olfactory bulb. *J. Neurosci.* **26**, 1247–1259 (2006).
14. Laurent, G. Olfactory network dynamics and the coding of multidimensional signals. *Nature Rev. Neurosci.* **3**, 884–895 (2002).
15. Willhite, D. C. *et al.* Viral tracing identified distributed columnar organization in the olfactory bulb. *Proc. Natl Acad. Sci. USA* **103**, 12592–12597 (2006).
16. Hildebrand, J. G. & Shepherd, G. M. Molecular mechanisms of olfactory discrimination: converging evidence for common principles across phyla. *Annu. Rev. Neurosci.* **20**, 595–631 (1997).
17. Shepherd, G. M. Outline of a theory of olfactory processing and its relevance to humans. *Chem. Senses* **30** (Suppl. 1), i3–i5 (2005).
18. Acree, T. E., Deibler, K. D. & Kittel, K. M. In *Handbook of Flavor Characterization: Sensory Analysis, Chemistry, and Physiology* (eds Deibler, K. D. & Delwiche, J.) 83–91 (Marcel Dekker, New York, 2004).
19. Rozin, P. 'Taste-smell confusions' and the duality of the olfactory sense. *Percept. Psychophys.* **31**, 397–401 (1982).
20. Sun, B. C. & Halpern, B. P. Identification of air phase retronasal and orthonasal odorant pairs. *Chem. Senses* **30**, 693–706 (2005).
21. Duffy, V. B., Backstrand, J. & Ferris, A. M. Olfactory dysfunction and related nutritional risk in free-living, elderly women. *J. Am. Diet. Assoc.* **95**, 879–884 (1995).
22. Shepherd-Barr, K. & Shepherd, G. M. Madeleines and neuromodernism: reassessing mechanisms of autobiographical memory in Proust. *Auto/Biography Studies* **13**, 40–60 (1998).
23. Taylor, A. J., Linforth, R. S. T., Harvey, B. A. & Blake, A. Atmospheric pressure chemical ionisation for monitoring of volatile flavour release *in vivo*. *Food Chem.* **71**, 327–338 (2000).
24. Bartoshuk, L. M. *et al.* From psychophysics to the clinic: missteps and advances. *Food Qual. Preference* **15**, 617–632 (2004).
25. Murphy, C., Cain, W. S. & Bartoshuk, L. M. Mutual action of taste and olfaction. *Sens. Process.* **1**, 204–211 (1977).

26. Mozell, M., Smith, B., Smith, P., Sullivan, L. & Swender, P. Nasal chemoreception in flavor identification. *Arch. Otolaryngol.* **90**, 367–373 (1969).
27. Farb, P. & Armelagos, G. *Consuming Passions: The Anthropology of Eating* (Houghton Mifflin, Boston, 1980).
28. Scalera, G., Grigson, P. S. & Norgren, R. Gustatory functions, sodium appetite, and conditioned taste aversion survive excitotoxic lesions of the thalamic taste area. *Behav. Neurosci.* **111**, 633–645 (1997).
29. Steiner, J. E. Discussion paper: innate, discriminative human facial expressions to taste and smell stimulation. *Ann. NY Acad. Sci.* **237**, 229–233 (1974).
30. Small, D. M., Gerber, J. C., Mak, Y. E. & Hummel, T. Differential neural responses evoked by orthonasal versus retronasal odorant perception in humans. *Neuron* **47**, 593–605 (2005).
31. Ongur, D., Ferry, A. T. & Price, J. L. Architectonic subdivision of the human orbital and medial prefrontal cortex. *J. Comp. Neurol.* **460**, 425–449 (2003).
32. Rolls, E. T. Taste, olfactory, and food texture processing in the brain, and the control of food intake. *Physiol. Behav.* **85**, 45–56 (2005).
33. Small, D. M. *et al.* Experience-dependent neural integration of taste and smell in the human brain. *J. Neurophysiol.* **92**, 1892–1903 (2004).
34. Gottfried, J. A. & Zald, D. H. On the scent of human olfactory orbitofrontal cortex: meta-analysis and comparison to non-human primates. *Brain Res. Brain Res. Rev.* **50**, 287–304 (2005).
35. Koza B. J., Cilimi, A., Dolese, M. & Zellner, D. A. Color enhances orthonasal olfactory intensity and reduces retronasal intensity. *Chem. Senses* **30**, 643–649 (2005).
36. Morrot, G., Brochet, F. & Dubourdieu, D. The color of odors. *Brain Lang.* **79**, 309–320 (2001).
37. Haberly, L. B. Parallel-distributed processing in olfactory cortex: new insights from morphological and physiological analysis of neuronal circuitry. *Chem. Senses* **26**, 551–576 (2001).
38. Wilson, D. A. & Stevenson, R. J. Olfactory perceptual learning: the critical role of memory in odor discrimination. *Neurosci. Biobehav.* **27**, 307–328 (2003).
39. Ross, R. S., McGaughy, J. & Eichenbaum, H. Acetylcholine in the orbitofrontal cortex is necessary for the acquisition of a socially transmitted food preference. *Learn. Mem.* **12**, 302–306 (2005).
40. Linster, C., Maloney, M., Patil, M. & Hasselmo, M. E. Enhanced cholinergic suppression of previously strengthened synapses enables the formation of self-organized representations in olfactory cortex. *Neurobiol. Learn. Mem.* **80**, 302–314 (2003).
41. Gietzen, D. W. & Rogers, Q. R. Nutritional homeostasis and indispensable amino acid sensing: a new solution to an old puzzle. *Trends Neurosci.* **29**, 91–99 (2006).
42. Rolls, E. T., Kringelbach, M. L. & de Araujo, I. E. Different representations of pleasant and unpleasant odours in the human brain. *Eur. J. Neurosci.* **18**, 695–703 (2003).
43. Pelchat, M. L., Johnson, A., Chan, R., Valdez, J. & Ragland, J. D. Images of desire: food-craving activation during fMRI. *Neuroimage* **23**, 1486–1493 (2004).
44. Volkow, N. D. & Wise, R. A. How can drug addiction help us understand obesity? *Nature Neurosci.* **8**, 555–560 (2005).
45. Pelchat, M. L. Of human bondage: food craving, obsession, compulsion and addiction. *Physiol. Behav.* **76**, 347–352 (2002).
46. Carpino, S. *et al.* Composition and aroma compounds of Ragusano cheese: native pasture and total mixed rations. *J. Dairy Sci.* **87**, 816–830 (2004).
47. Deibler, K. D. & Delwiche, J. (eds) *Handbook of Flavor Characterization: Sensory Analysis, Chemistry, and Physiology* (Marcel Dekker, New York, 2003).
48. Licita, G. (ed.) *Cheese Art 2004. Ragusa: Consorzio Ricerca Filiera Lattiero Casearia*. (Ragusa, Italy, 5 Jun 2004).
49. Duffy, V. B., Cain, W. S. & Ferris, A. M. Measurement of sensitivity to olfactory flavor: application in a study of aging and dentures. *Chem. Senses* **24**, 671–677 (1999).
50. Corriher, S. O. *CookWise* (William Morrow, New York, 1997).
51. McGee, H. *On Food and Cooking: The Science and Lore of the Kitchen* (Scribner, New York, 2004).
52. This, H. Molecular gastronomy. *Nature Mater.* **4**, 5–7 (2005).
53. Mori, K., Nagao, H. & Yoshihara, Y. The olfactory bulb: coding and processing of odor molecule information. *Science* **286**, 711–715 (1999).
54. Smith, D. V. & Frank, M. E. in *Mechanisms of Taste Transduction* (eds Simon, S. A. & Roper, S. D.) 295–338 (CRC Press, Baton Rouge, 1993).
55. Rolls, E. T. The representation of information about faces in the temporal and frontal lobes. *Neuropsychologia* [Epub ahead of print] 22 Jun 2006 (doi:10.1016/j.neuropsychologia.2006.04.019).
56. Lorig, T. S. On the similarity of odor and language perception. *Neurosci. Biobehav. Rev.* **23**, 391–398 (1999).
57. Herz, R. S. The effect of verbal context on olfactory perception. *J. Exp. Psychol. Gen.* **132**, 595–606 (2003).
58. Wrangham, R. & Conklin-Brittain, N. Cooking as a biological trait. *Comp. Biochem. Physiol. A Mol. Integr. Physiol.* **136**, 35–46 (2003).

Acknowledgements Research in my laboratory is supported by the National Institutes of Health National Institute for Deafness and Other Communicative Disorders, and by the Human Brain Project. I thank M. L. Pelchat, E. T. Rolls, D. Small, L. Bartoshuk, T. Acree, F. Q. Xu and V. Duffy for valuable advice.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The author declares no competing financial interests. Correspondence should be addressed to the author (gordon.shepherd@yale.edu).

The plant immune system

Jonathan D. G. Jones¹ & Jeffery L. Dangl²

Many plant-associated microbes are pathogens that impair plant growth and reproduction. Plants respond to infection using a two-branched innate immune system. The first branch recognizes and responds to molecules common to many classes of microbes, including non-pathogens. The second responds to pathogen virulence factors, either directly or through their effects on host targets. These plant immune systems, and the pathogen molecules to which they respond, provide extraordinary insights into molecular recognition, cell biology and evolution across biological kingdoms. A detailed understanding of plant immune function will underpin crop improvement for food, fibre and biofuels production.

Introduction

Plant pathogens use diverse life strategies. Pathogenic bacteria proliferate in intercellular spaces (the apoplast) after entering through gas or water pores (stomata and hydathodes, respectively), or gain access via wounds. Nematodes and aphids feed by inserting a stylet directly into a plant cell. Fungi can directly enter plant epidermal cells, or extend hyphae on top of, between, or through plant cells. Pathogenic and symbiotic fungi and oomycetes can invaginate feeding structures (haustoria), into the host cell plasma membrane. Haustorial plasma membranes, the extracellular matrix, and host plasma membranes form an intimate interface at which the outcome of the interaction is determined. These diverse pathogen classes all deliver effector molecules (virulence factors) into the plant cell to enhance microbial fitness.

Plants, unlike mammals, lack mobile defender cells and a somatic adaptive immune system. Instead, they rely on the innate immunity of each cell and on systemic signals emanating from infection sites^{1–3}. We previously reviewed disease resistance (*R*) protein diversity, polymorphism at *R* loci in wild plants and lack thereof in crops, and the suite of cellular responses that follow *R* protein activation¹. We hypothesized that many plant *R* proteins might be activated indirectly by pathogen-encoded effectors, and not by direct recognition. This ‘guard hypothesis’ implies that *R* proteins indirectly recognize pathogen effectors by monitoring the integrity of host cellular targets of effector action^{1,4}. The concept that *R* proteins recognize ‘pathogen-induced modified self’ is similar to the recognition of ‘modified self’ in ‘danger signal’ models of the mammalian immune system⁵.

It is now clear that there are, in essence, two branches of the plant immune system. One uses transmembrane pattern recognition receptors (PRRs) that respond to slowly evolving microbial- or pathogen-associated molecular patterns (MAMPs or PAMPs), such as flagellin⁶. The second acts largely inside the cell, using the polymorphic NB-LRR protein products encoded by most *R* genes¹. They are named after their characteristic nucleotide binding (NB) and leucine rich repeat (LRR) domains. NB-LRR proteins are broadly related to animal CATERPILLER/NOD/NLR proteins⁷ and STAND ATPases⁸. Pathogen effectors from diverse kingdoms are recognized by NB-LRR proteins, and activate similar defence responses. NB-LRR-mediated disease resistance is effective against pathogens that can grow only on living host tissue (obligate biotrophs), or hemibiotrophic pathogens, but not against pathogens that kill host tissue during colonization (necrotrophs)⁹.

Our current view of the plant immune system can be represented as a four phased ‘zigzag’ model (Fig. 1), in which we introduce several

important abbreviations. In phase 1, PAMPs (or MAMPs) are recognized by PRRs, resulting in PAMP-triggered immunity (PTI) that can halt further colonization. In phase 2, successful pathogens deploy effectors that contribute to pathogen virulence. Effectors can interfere with PTI. This results in effector-triggered susceptibility (ETS). In phase 3, a given effector is ‘specifically recognized’ by one of the NB-LRR proteins, resulting in effector-triggered immunity (ETI). Recognition is either indirect, or through direct NB-LRR recognition of an effector. ETI is an accelerated and amplified PTI response, resulting in disease resistance and, usually, a hypersensitive cell death response (HR) at the infection site. In phase 4, natural selection drives pathogens to avoid ETI either by shedding or diversifying the recognized effector gene, or by acquiring additional effectors that suppress ETI. Natural selection results in new *R* specificities so that

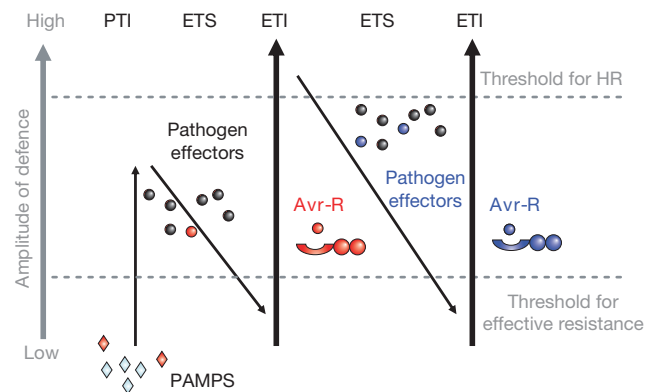


Figure 1 | A zigzag model illustrates the quantitative output of the plant immune system. In this scheme, the ultimate amplitude of disease resistance or susceptibility is proportional to [PTI – ETS + ETI]. In phase 1, plants detect microbial/pathogen-associated molecular patterns (MAMPs/PAMPs, red diamonds) via PRRs to trigger PAMP-triggered immunity (PTI). In phase 2, successful pathogens deliver effectors that interfere with PTI, or otherwise enable pathogen nutrition and dispersal, resulting in effector-triggered susceptibility (ETS). In phase 3, one effector (indicated in red) is recognized by an NB-LRR protein, activating effector-triggered immunity (ETI), an amplified version of PTI that often passes a threshold for induction of hypersensitive cell death (HR). In phase 4, pathogen isolates are selected that have lost the red effector, and perhaps gained new effectors through horizontal gene flow (in blue)—these can help pathogens to suppress ETI. Selection favours new plant NB-LRR alleles that can recognize one of the newly acquired effectors, resulting again in ETI.

¹The Sainsbury Laboratory, John Innes Centre, Norwich Research Park, Colney, Norwich NR4 7UH, UK. ²Department of Biology, Department of Microbiology and Immunology, Curriculum in Genetics, and Carolina Center for Genome Sciences, CB-3280, University of North Carolina, Chapel Hill, North Carolina 27599, USA.

ETI can be triggered again. Below, we review each phase in turn, we update the experimental validation of the 'guard hypothesis', and we consider future challenges in understanding and manipulating the plant immune system. We will not discuss the small RNA-based plant immune system active against viruses¹⁰ or the active response of plants to herbivores¹¹.

Microbial patterns and plant pattern recognition

We define basal disease resistance as that activated by virulent pathogens on susceptible hosts. Thus, basal disease resistance is, at first glance, PTI minus the effects of ETs; however, there is also likely to be weak ETI triggered by weak recognition of effectors, as detailed below. Hence, the most accurate definition of basal defence would be 'PTI plus weak ETI, minus ETs'. The archetypal elicitor of PTI is bacterial flagellin, which triggers defence responses in various plants¹². Flagellum-based motility is important for bacterial pathogenicity in plants⁶. A synthetic 22-amino-acid peptide (flg22) from a conserved flagellin domain is sufficient to induce many cellular responses¹³ including the rapid (<1 h) transcriptional induction of at least 1,100 *Arabidopsis thaliana* (hereafter *Arabidopsis*) genes¹⁴. A genetic screen using flg22 defined the *Arabidopsis* LRR-receptor kinase FLS2, which binds flg22 (ref. 15). FLS2 and mammalian TLR5 recognize different flagellin domains⁶. FLS2 is internalized following stimulation by a receptor-mediated endocytic process that presumably has regulatory functions¹⁶. *fls2* mutants exhibit enhanced sensitivity to spray application of pathogenic *Pseudomonas syringae* pv. *tomato* DC3000 (*Pto* DC3000), but not to syringe infiltration into the leaf apoplast^{14,17}, suggesting that FLS2 acts early against pathogen invasion.

Bacterial cold shock proteins and elongation factor Tu (EF-Tu) activate similar defence responses to flg22 (refs 18–20). Ef-Tu is recognized by an *Arabidopsis* LRR-kinase called EFR (ref. 20). *efr* mutants support higher levels of transient transformation with *Agrobacterium*, suggesting that PTI might normally limit *Agrobacterium* pathogenicity. Treatment with a conserved EF-Tu peptide induces expression of a gene set nearly identical to that induced by flg22 (ref. 20). Conversely, EFR transcription is induced by flg22. Hence, the responses to MAMPs/PAMPs converge on a limited number of signalling pathways and lead to a common set of outputs that comprise PTI. Remarkably, mutations in genes required for NB-LRR function have no effect on early responses to flg22 (ref. 14). Thus, NB-LRR-dependent signalling and MAMP/PAMP-mediated signalling require partially distinct components.

Molecules that induce PTI are not easily discarded by the microbes that express them. Yet flagellin from various *Xanthomonas campestris* pv. *campestris* strains is variably effective in triggering FLS2-mediated PTI in *Arabidopsis*, and flagellin from *Agrobacterium tumefaciens* or *Sinorhizobium meliloti* is less active than that from *P. syringae*¹³. Ef-Tu from *Pto* DC3000 is much less active in eliciting PTI in *Arabidopsis* than is Ef-Tu from *Agrobacterium*¹⁹. Limited variation also exists in PAMP responsiveness within a plant species. *Arabidopsis* accession Ws-0 carries a point mutation in FLS2, rendering it non-responsive to flg22 (ref. 12). In fact, individual plant species recognize only a subset of potential PAMPs (ref. 6). Neither PAMPs nor PRRs are invariant, and each can be subject to natural selection.

Additional MAMPs/PAMPs and corresponding PRRs must exist, since *Agrobacterium* extracts elicit PTI on an *fls2 efr-1* double mutant (ref. 20). Other LRR kinases may encode additional PRRs whose transcription is stimulated by engagement of related PRRs. There are over 200 LRR-kinases in the *Arabidopsis* Col-0 genome²¹; 28 of these are induced within 30 min of flg22 treatment¹⁴. FLS2 and EFR are members of an atypical kinase family that might have a plant immune system specific function²². There are also 56 *Arabidopsis* receptor-like proteins (RLPs) that encode type I transmembrane proteins with LRR ectodomains, but no intracellular kinase domains²³. MAMP/PAMP elicitation might 'prime' further defence responses by elevating responsiveness to other microbial patterns¹⁴.

Successful pathogens suppress PTI

What does a would-be pathogen, using its collection of effectors, need to achieve? Some effectors may serve structural roles, for example, in the extrahaustorial matrix that forms during fungal and oomycete infection²⁴. Others may promote nutrient leakage or pathogen dispersal²⁵. Many are likely to contribute to suppression of one or more components of PTI or ETI. The extent to which ETI and PTI involve distinct mechanisms is still an open question, and some effectors may target ETI rather than PTI, or vice versa (Fig. 1).

Plant pathogenic bacteria deliver 15–30 effectors per strain into host cells using type III secretion systems (TTSS). Bacterial effectors contribute to pathogen virulence, often by mimicking or inhibiting eukaryotic cellular functions^{26–28}. A pathogenic *P. syringae* strain mutated in the TTSS, and unable to deliver any type III effectors, triggers a faster and stronger transcriptional re-programming in bean than does the isogenic wild-type strain²⁹. This strain, representing the sum of all bacterial MAMPs/PAMPs, induces transcription of essentially the same genes as flg22 (refs 30–32). Hence, the type III effectors from any successful bacterial pathogen dampen PTI sufficiently to allow successful colonization³³ (Fig. 1).

Excellent reviews discuss cellular processes targeted by bacterial type III effectors^{26–28,34}; we highlight only new examples. The *P. syringae* HopM effector targets at least one ARF-GEF protein likely to be involved in host cell vesicle transport³⁵. HopM functions redundantly with the unrelated effector AvrE in *P. syringae* virulence³⁶ suggesting that manipulation of host vesicle transport is important for successful bacterial colonization. AvrPto and AvrPtoB are unrelated type III effectors that may contribute to virulence by inhibiting early steps in PTI, upstream of MAPKKK (ref. 37). Like other type III effectors, AvrPtoB is a bipartite protein. The amino terminus contributes to virulence; the carboxy terminus may have a function in blocking host cell death^{38,39}. A domain from the AvrPtoB C terminus folds into an active E3 ligase, suggesting that its function involves host protein degradation⁴⁰. The *Yersinia* effector YopJ, a member of the AvrRxv family of effectors from phytopathogenic bacteria, inhibits MAP kinase cascades by acetylation of phosphorylation-regulated residues on a MEK protein⁴¹. Many additional bacterial type III effectors protein families have been identified^{26–28}; their targets and functions await definition.

Effectors from plant pathogens that are eukaryotic are poorly understood. Fungal and oomycete effectors can act either in the extracellular matrix or inside the host cell. For example, the tomato RLPs, Cf-2, Cf-4, Cf-5 and Cf-9 respond specifically to extracellular effectors produced by *Cladosporium fulvum*⁴². Other fungal and oomycete effectors probably act inside the host cell; they are recognized by NB-LRR proteins. For example, the gene encoding the oomycete effector Atr13 from *Hyaloperonospora parasitica* exhibits extensive allelic diversity between *H. parasitica* strains matched by diversity at the corresponding *Arabidopsis* RPP13 NB-LRR locus⁴³. Diversity is also observed across *H. parasitica* Atr1 and *Arabidopsis* RPP1 alleles⁴⁴. Atr1 and Atr13 carry signal peptides for secretion from *H. parasitica*. They share with each other, and with the *Phytophthora infestans* Avr3a protein, an RxLR motif, that enables import of *Plasmodium* effectors into mammalian host cells⁴⁵. This is consistent with the taxonomic proximity of oomycetes and *Plasmodium*. Races of the flax rust fungus *Melampsora lini* express Avr genes recognized by specific alleles of the flax L, M and P NB-LRR proteins. These haustorial proteins carry signal peptides for fungal export and can function inside the plant cell^{46,47}. How they are taken up by the host cell is unknown. However, the barley powdery mildew (*Blumeria graminis* f.sp. *hordei*) AvrK and AvrA10 proteins, recognized by the NB-LRR barley genes *Mlk* and *Mla10*, contain neither obvious signal peptides nor RxLR motifs, yet are members of large gene families in *Blumeria* and *Erysiphe* species⁴⁸. How these oomycete and fungal effectors are delivered to the host cell and contribute to pathogen virulence is unknown.

Pathogens produce small molecule effectors that mimic plant hormones. Some *P. syringae* strains make coronatine, a jasmonic acid mimic that suppresses salicylic-acid-mediated defence to biotrophic pathogens^{49,50} and induces stomatal opening, helping pathogenic bacteria gain access to the apoplast⁵¹. PTI involves repression of auxin responses, mediated in part by a micro-RNA that is also induced during abscisic-acid-mediated stress responses⁵². Gibberellin is produced by the fungal pathogen *Gibberella fujikuroi* leading to 'foolish seedling' syndrome, and cytokinin produced by many pathogens can promote pathogen success through retardation of senescence in infected leaf tissue. The interplay between PTI and normal hormone signalling, and pathogen mimics that influence it, is just beginning to be unravelled.

Indirect and direct host recognition of pathogen effectors

Effectors that enable pathogens to overcome PTI are recognized by specific disease resistance (*R*) genes. Most *R* genes encode NB-LRR proteins; there are ~125 in the *Arabidopsis* Col-0 genome. If one effector is recognized by a corresponding NB-LRR protein, ETI ensues. The recognized effector is termed an avirulence (Avr) protein. ETI is a faster and stronger version of PTI^{30–32} that often culminates in HR⁵³ (Fig. 1). HR typically does not extend beyond the infected cell: it may retard pathogen growth in some interactions, particularly those involving haustorial parasites, but is not always observed, nor required, for ETI. It is unclear what actually stops pathogen growth in most cases.

Very little is known about the signalling events required to activate NB-LRR-mediated ETI. NB-LRR proteins are probably folded in a signal competent state by cytosolic heat shock protein 90 and other receptor co-chaperones^{54,55}. The LRRs seem to act as negative regulators that block inappropriate NB activation. NB-LRR activation involves intra- and intermolecular conformational changes and may resemble the induced proximity mechanism by which the related animal Apaf-1 protein activates programmed cell death⁵⁶.

NB-LRR activation results in a network of cross-talk between response pathways deployed, in part, to differentiate biotrophic from necrotrophic pathogen attack⁹. This is maintained by the balance between salicylic acid, a local and systemic signal for resistance against many biotrophs, and the combination of jasmonic acid and ethylene accumulation as signals that promote defence against necrotrophs⁹. Additional plant hormones are likely to alter the salicylic-acid-jasmonic-acid/ethylene signalling balance. *Arabidopsis* mutants defective in salicylic acid biosynthesis or responsiveness are compromised in both basal defence and systemic acquired resistance (SAR)⁵⁷. NB-LRR activation induces differential salicylic-acid- and ROS-dependent responses at and surrounding infection sites, and systemically⁵⁸. The NADPH-oxidase-dependent oxidative burst that accompanies ETI represses salicylic acid-dependent cell death spread in cells surrounding infection sites⁵⁹. Local and systemic changes in gene expression are mediated largely by transcription factors of the WRKY and TGA families⁶⁰.

Several NB-LRR proteins recognize type III effectors indirectly, by detecting products of their action on host targets, consistent with the 'guard hypothesis'¹. The key tenets of this hypothesis are that: (1) an effector acting as a virulence factor has a target(s) in the host; (2) by manipulating or altering this target(s) the effector contributes to pathogen success in susceptible host genotypes; and (3) effector perturbation of a host target generates a 'pathogen-induced modified-self' molecular pattern, which activates the corresponding NB-LRR protein, leading to ETI. Three important consequences of this model, now supported by experimental evidence, are that: (1) multiple effectors could evolve independently to manipulate the same host target, (2) this could drive the evolution of more than one NB-LRR protein associated with a target of multiple effectors, and (3) these NB-LRRs would be activated by recognition of different modified-self patterns produced on the same target by the action of the effectors in (1).

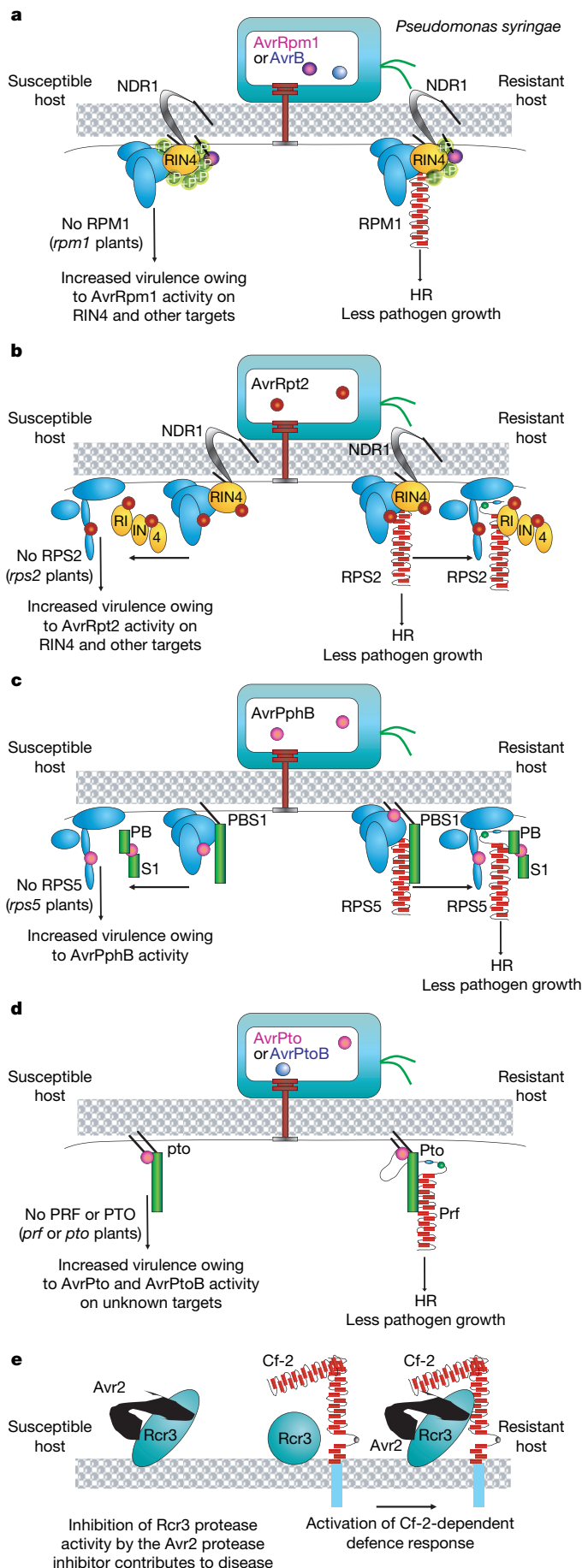
RIN4, a 211-amino-acid, acylated⁶¹ and plasma-membrane-associated protein, is an archetypal example of a host target of type III effectors that is guarded by NB-LRR proteins (Fig. 2). It is manipulated by three different bacterial effectors, and associates *in vivo* with two *Arabidopsis* NB-LRR proteins (Fig. 2a and 2b). Two unrelated type III effectors, AvrRpm1 and AvrB, interact with and induce phosphorylation of RIN4 (ref. 62). This RIN4 modification is predicted to activate the RPM1 NB-LRR protein. A third effector, AvrRpt2 is a cysteine protease⁶³, activated inside the host cell⁶⁴, that eliminates RIN4 by cleaving it at two sites^{61,65}. Cleavage of RIN4 activates the RPS2 NB-LRR protein^{66,67}. Activation of both RPM1 and RPS2 requires the GPI-anchored NDR1 protein, and RIN4 interacts with NDR1⁶⁸.

If RIN4 was the only target for these three effectors, then its elimination would abolish their ability to add virulence to a weakly pathogenic strain. However, elimination of RIN4 demonstrated that it is not the only host target for AvrRpm1 or AvrRpt2 in susceptible (*rin4 rpm1 rps2*) plants⁶⁹. Additionally, AvrRpt2 can cleave *in vitro* several *Arabidopsis* proteins that contain its consensus cleavage site⁶⁵. Hence, any effector's contribution to virulence might involve manipulation of several host targets, and the generation of several modified-self molecules. However, the perturbation of only one target is sufficient for NB-LRR activation. RIN4 negatively regulates RPS2 and RPM1 (and only these two NB-LRR proteins)^{69,70}. But what is the function of RIN4 in the absence of RPS2 and RPM1? In *rpm1 rps2* plants, AvrRpt2 or AvrRpm1 (and possibly other effectors) manipulate RIN4 (and possibly associated proteins or other targets) in order to suppress PTI⁷¹. Thus, plants use NB-LRR proteins to guard against pathogens that deploy effectors to inhibit PAMP-signalling. Additional examples of indirect recognition are detailed in Fig. 2; these include both intra- and extra-cellular recognition of pathogen-induced modified self.

Not all NB-LRR recognition is indirect, and there are three examples of direct Avr-NB-LRR interaction^{72–74}. The flax *L* locus alleles encode NB-LRR proteins that interact in yeast with the corresponding AvrL proteins, providing the first evidence that effector diversity determining NB-LRR recognition can be correlated perfectly with effector-NB-LRR-protein interaction⁷². Both *L* and AvrL proteins are under diversifying selection, arguing for a direct evolutionary arms race. The allelic diversity of other fungal and oomycete pathogen effectors, and of their corresponding host NB-LRR proteins as described above, also suggests direct interaction, though this remains to be demonstrated.

The evolutionary radiation of several hundred thousand angiosperm plant species ~140–180 million years ago was probably accompanied by many independent cases of pathogen co-evolution, particularly of host-adapted obligate biotrophs. Most plants resist infection by most pathogens; they are said to be 'non-hosts'. This non-host resistance could be mediated by at least two mechanisms. First, a pathogen's effectors could be ineffective on a potential new, but evolutionarily divergent, host, resulting in little or no suppression of PTI, and failure of pathogen growth. Alternatively, one or more of the effector complement of the would-be pathogen could be recognized by the NB-LRR repertoire of plants other than its co-adapted host, resulting in ETI. These two scenarios predict different outcomes with respect to the timing and amplitude of the response they would trigger, and they also give rise to different evolutionary pressures on both host and pathogen.

Non-host resistance in *Arabidopsis* against the non-adapted barley pathogen, *B. graminis* f. sp. *hordei* (Bgh) normally involves the rapid production of cell wall appositions (physical barriers) and antimicrobial metabolites at the site of pathogen entry, but no HR. *Arabidopsis* penetration (*pen*) mutants are partially compromised in this response. PEN2 is a peroxisomal glucosyl hydrolase⁷⁵, and PEN3 encodes a plasma membrane ABC transporter⁷⁶. PEN2 and PEN3 are both recruited to attempted fungal entry sites, apparently to mediate the polarized delivery of a toxin to the apoplast^{75,76}. The



actin cytoskeleton probably contributes to this response⁷⁷, perhaps as a track for PEN2-containing peroxisomes and/or vesicles. This pre-invasion non-host resistance is genetically separable from a post-invasion mechanism that requires additional elements that regulate both PTI and ETI⁷⁸. Elimination of both PEN2 and PTI/ETI signalling transforms *Arabidopsis* into a host for an evolutionarily non-adapted fungal pathogen⁷⁵. This suggests that non-host resistance comprises mechanistically distinct layers of resistance.

The PEN1 syntaxin acts in a different pre-invasion non-host resistance pathway. PEN1 is likely to be part of a ternary SNARE complex that secretes vesicle cargo to the site of attempted fungal invasion, contributing to formation of cell wall appositions^{79–81}. Specific seven-transmembrane MLO (mildew resistance locus O) family members negatively regulate PEN1-dependent secretion at sites of attempted pathogen ingress^{79,80}. Recessive *mlo* mutations in either *Arabidopsis* or barley result in resistance to the respective co-evolved powdery mildew pathogens⁸². Hence, in both *Arabidopsis* and barley, these fungi might suppress PEN1-mediated disease resistance by activation of MLO. This remarkable set of findings implies that a common host cell entry mechanism evolved in powdery mildew fungi at or before the monocot–dicot divergence. *PEN2* and *PEN3* genes are induced by flg22, indicating that they might be involved in PTI.

Non-host resistance can also be mediated by parallel ETI responses. For example, four bacterial effectors from a tomato pathogen

Figure 2 | Plant immune system activation by pathogen effectors that generate modified self molecular patterns.

a, *Arabidopsis* RPM1 is a peripheral plasma membrane NB-LRR protein. It is activated by either the AvrRpm1 or the AvrB effector proteins. AvrRpm1 enhances the virulence of some *P. syringae* strains on *Arabidopsis* as does AvrB on soybeans. AvrRpm1 and AvrB are modified by eukaryote-specific acylation once delivered into the cell by the type III secretion system (red syringe) and are thus targeted to the plasma membrane. The biochemical functions of AvrRpm1 and AvrB are unknown, although they target RIN4, which becomes phosphorylated (+P), and activate RPM1, as detailed in the text. In the absence of RPM1, AvrRpm1 and AvrB presumably act on RIN4 and other targets to contribute to virulence. Light blue eggs in this and subsequent panels represent as yet unknown proteins. **b**, RPS2 is an NB-LRR protein that resides at the plasma membrane. It is activated by the AvrRpt2 cysteine protease type III effector from *P. syringae*. Auto-processing of AvrRpt2 by a host cyclophilin reveals a consensus, but unconfirmed, myristoylation site at the new amino terminus, suggesting that it might also be localized to the host plasma membrane. AvrRpt2 is the third effector that targets RIN4. Cleavage of RIN4 by AvrRpt2 leads to RPS2-mediated ETI. In the absence of RPS2, AvrRpt2 presumably cleaves RIN4 and other targets as part of its virulence function. **c**, RPS5 is an *Arabidopsis* NB-LRR protein localized to a membrane fraction, probably via acylation. RPS5 is NDR1-independent. It is activated by the AvrPphB cysteine protease effector from *P. syringae*¹⁰⁰. AvrPphB is cleaved, acylated and delivered to the host plasma membrane. Activated AvrPphB cleaves the *Arabidopsis* PBS1 serine-threonine protein kinase, leading to RPS5 activation. The catalytic activity of cleaved PBS1 is required for RPS5 activation, suggesting that this ‘modified-self’ fragment retains its enzymatic activity as part of the RPS5 activation mechanism¹⁰⁰. To date, no function has been ascribed to PBS1 in the absence of RPS5. **d**, Pto is a tomato serine-threonine protein kinase. Pto is polymorphic and hence satisfies the genetic criteria for the definition of a disease resistance protein. Pto activity requires the NB-LRR protein Prf, and the proteins form a molecular complex¹⁰¹. Prf is monomorphic, at least in the tomato species analysed to date. Pto is the direct target of two unrelated *P. syringae* effectors, AvrPto and AvrPtoB, each of which contributes to pathogen virulence in *pto* mutants¹⁰². It is thus likely that Prf guards Pto (refs 101, 103). The Pto kinase is apparently not required for PTI, though there may be redundancy in its function because it is a member of a gene family. **e**, The transmembrane RLP Cf-2 guards the extracellular cysteine protease Rcr3. Cf-2 recognizes the *C. fulvum* extracellular effector Avr2, which encodes a cysteine protease inhibitor. Avr2 binds and inhibits the tomato Rcr3 cysteine protease. Mutations in Rcr3 result in the specific loss of Cf-2-dependent recognition of Avr2. Hence, Cf-2 seems to monitor the state of Rcr3, and activates defence if Rcr3 is inhibited by Avr2 (ref. 104).

unable to colonize soybean can each trigger specific soybean *R* genes when delivered from a soybean pathogen⁸³. Deletion of these effector genes from the tomato pathogen diminishes its virulence on tomato, but does not allow it to colonize soybean⁸⁴. Hence, there might be other factors lacking in this strain that are required to colonize soybean. Also, a widely distributed, monomorphic effector acting as an avirulence protein is sufficient to render *Magnaporthe oryzae* strains unable to colonize rice. Its presence in over 50 strains that successfully colonize perennial ryegrass suggests a virulence function⁸⁵. Finally, *Arabidopsis* non-host resistance to *Leptosphaeria maculans*, a fungal pathogen of *Brassica*, is actually mediated by unlinked NB-LRR proteins present in each parent of a cross between two accessions⁸⁶. Hence, cryptic NB-LRR mediated responses acting in parallel can limit pathogen host range.

Pathogens dodge host surveillance

The effectiveness of ETI selects for microbial variants that can avoid NB-LRR-mediated recognition of a particular effector (Fig. 3). Effector allele frequencies are likely to be influenced by their mode of action. The diversity of both flax rust AvrL alleles and oomycete Atr13 and Atr1 alleles suggests one means of effector evolution. These proteins are likely to interact directly *in planta* with proteins encoded by alleles of the flax *L* and *Arabidopsis* *RPP1* and *RPP13* loci, respectively. The high level of diversifying selection among these effector alleles is presumably selected by host recognition, and hence acts on effector residues that are probably not required for effector function.

In contrast, effectors providing biochemical functions that generate modifications of host targets are likely to be under purifying selection⁸⁷. NB-LRR activation via recognition of pathogen-induced modified self provides a mechanism for host perception of multiple effectors evolved to compromise the same host target (Fig. 2). For selection to generate an effector that escapes ETI, the effector is likely to lose its nominal function. The simplest pathogen response to host recognition is to jettison the detected effector gene, provided the population's effector repertoire can cover the potential loss of fitness

on susceptible hosts. In fact, effector genes are often associated with mobile genetic elements or telomeres and are commonly observed as presence/absence polymorphisms across bacterial and fungal strains. Indirect recognition of effector action, via recognition of pathogen-induced modified self, is likely to enable relatively stable, durable and evolutionarily economic protection of the set of cellular machinery targeted by pathogen effectors.

ETI can also be overcome through evolution of pathogen effectors that suppress it directly (Fig. 1). For example, in *P. syringae* pv. *phaseolicola*, the AvrPphC effector suppresses ETI triggered by the AvrPphF effector in some cultivars of bean, whereas, as its name implies, AvrPphC itself can condition avirulence on different bean cultivars⁸⁸. Other cases of bacterial effectors acting to dampen or inhibit ETI have been observed³⁸. Genetic analysis in flax rust revealed so-called inhibitor genes that function to suppress ETI triggered by other avirulence genes⁸⁹. Hence, it seems likely that some effectors suppress the ETI triggered by other effectors.

Microbial evolution in response to ETI may result in two extremes of NB-LRR evolution. NB-LRR gene homologues in diverse *Arabidopsis* accessions accumulate evolutionary novelty at different rates at different loci⁹⁰. Some NB-LRR genes are not prone to duplication, and are evolving relatively slowly. Their products are perhaps stably associated with a host protein whose integrity they monitor, retarding diversification. Others are evolving more rapidly and may interact directly with rapidly evolving effectors^{91,92}. What might drive these evolutionary modes (Fig. 3)? In pathogen populations, the frequency of an effector gene will be enhanced by its ability to promote virulence, and reduced by host recognition. For example, in the flax/flax rust system, avirulence gene frequency in the rust is elevated on plant populations with a lower abundance of the corresponding *R* genes, consistent with *avr* genes increasing pathogen fitness⁹³. Hence, natural selection should maintain effector function in the absence of recognition. But effector function has a cost that is dependent on the frequency of the corresponding *R* gene. And *R* genes may exact a fitness cost in the host⁹⁴. Thus, if effector frequency drops in a pathogen population, hosts might be selected for loss of the corresponding *R* allele, and the frequency-dependent cycle would continue (Fig. 3).

Challenges and opportunities for the future

We need to define the repertoire and modes of action of effectors from pathogens with diverse life histories. This will help to define the comprehensive set of host targets, as well as the evolutionary pressures acting on both hosts and pathogens. For example, if the majority of bacterial effectors evolve under purifying selection to maintain intrinsic function⁸⁷, then a population-wide set of unrelated microbial effectors might converge onto a limited set of NB-LRR-associated host targets. These stable associations of NB-LRR proteins and the host proteins whose integrity they monitor are presumably being challenged by newly evolved, or newly acquired, effectors that can still surreptitiously manipulate the target in the service of virulence.

We need to understand the haustorial interface. Rewiring of host and microbe vesicle traffic will probably underpin haustorial differentiation. We do not yet know whether the extrahaustorial membrane is derived from the host plasma membrane or is a newly synthesized, novel host membrane. High-throughput sequencing renders it feasible to index the gene complements of obligate biotrophs like powdery and downy mildews, and rusts. Genomics can be also be used to identify the genes expressed by the pathogen over a time course of infection. The presence of a signal peptide and (in the case of oomycetes) an RxLR motif, can then be used to computationally identify the complement of effector candidates. With the development of appropriate high-throughput delivery systems, it will be possible to investigate their functions, and their ability to impinge on PTI and/or ETI, on both host plants and other plant species.

Do the transcriptional controls of PTI and ETI, which culminate in similar outputs, overlap? Several effectors can be nuclear localized^{95,96}.

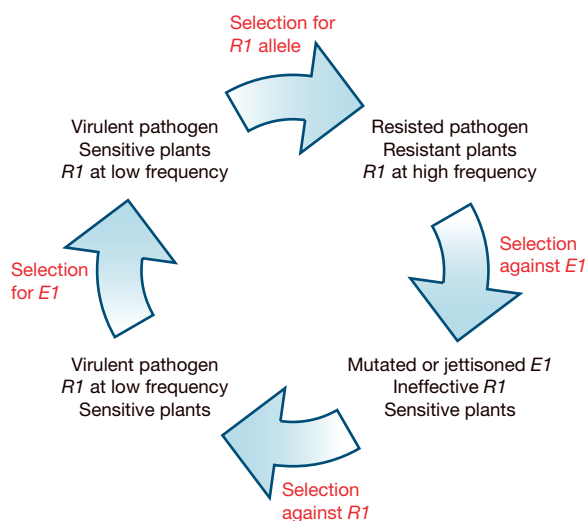


Figure 3 | Co-evolution of host *R* genes and the pathogen effector complement. A pathogen carries an effector gene (*E1*) that is recognized by a rare *R1* allele (top). This results in selection for an elevated frequency of *R1* in the population. Pathogens in which the effector is mutated are then selected, because they can grow on *R1*-containing plants (right). *R1* effectiveness erodes, and, because at least some *R* genes have associated fitness costs⁹⁴, plants carrying *R1* can have reduced fitness (bottom), resulting in reduced *R1* frequencies. The pathogen population will still contain individuals with *E1*. In the absence of *R1*, *E1* will confer increased fitness, and its frequency in the population will increase (left). This will lead to resumption of selection for *R1* (top). In populations of plants and pathogens, this cycle is continuously turning, with scores of effectors and many alleles at various *R* loci in play.

The RRS1 R protein is perhaps a Rosetta Stone chimaera of NB-LRR and WRKY transcription factor⁷⁴. A TATA-binding protein accessory factor, AtTIP49a, interacts with RPM1 and other NB-LRR proteins, and is a negative regulator of defence⁹⁷. Two nucleoporins and importins are required for the output response of an ectopically activated NB-LRR protein^{98,99}. Notably, the prototypic animal CATERPILLAR protein is CIITA, a transcriptional co-activator of MHC class II response to viral infection that is, in turn, the target of viral proteins that aim to shut it down⁷. Whether NB-LRR proteins are transcriptional co-regulators is at present unknown.

We need to know what causes pathogen growth arrest. Because plants are sessile, they must continuously integrate both biotic and abiotic signals from the environment. Plants lack circulating cells, so these responses also need to be partitioned both locally over several cell diameters and systemically over metres. Understanding the spatial interplay of PTI, ETI, and plant hormone and abiotic stress-signalling systems is in its infancy.

Finally, we need to understand the population biology of pathogen effectors, and their co-evolving host NB-LRR genes. Knowledge of their allele frequencies and their spatial distribution in wild ecosystems should tell us more about the evolution of this fascinating ancient immune system and how we might deploy it more effectively to control disease.

- Dangl, J. L. & Jones, J. D. G. Plant pathogens and integrated defence responses to infection. *Nature* **411**, 826–833 (2001).
- Ausubel, F. M. Are innate immune signaling pathways in plants and animals conserved? *Nature Immunol.* **6**, 973–979 (2005).
- Chisholm, S. T., Coaker, G., Day, B. & Staskawicz, B. J. Host–microbe interactions: shaping the evolution of the plant immune response. *Cell* **124**, 803–814 (2006).
- van der Biezen, E. A. & Jones, J. D. G. Plant disease resistance proteins and the gene-for-gene concept. *Trends Biochem. Sci.* **23**, 454–456 (1998).
- Matzinger, P. The danger model: a renewed sense of self. *Science* **296**, 301–305 (2002).
- Zipfel, C. & Felix, G. Plants and animals: a different taste for microbes? *Curr. Opin. Plant Biol.* **8**, 353–360 (2005).
- Ting, J. P. & Davis, B. K. CATERPILLER: a novel gene family important in immunity, cell death, and diseases. *Annu. Rev. Immunol.* **23**, 387–414 (2005).
- Leipe, D. D., Koonin, E. V. & Aravind, L. STAND, a class of P-loop NTPases including animal and plant regulators of programmed cell death: multiple, complex domain architectures, unusual phylogenetic patterns, and evolution by horizontal gene transfer. *J. Mol. Biol.* **343**, 1–28 (2004).
- Glazebrook, J. Contrasting mechanisms of defense against biotrophic and necrotrophic pathogens. *Annu. Rev. Phytopathol.* **43**, 205–227 (2005).
- Voinnet, O. Induction and suppression of RNA silencing: insights from viral infections. *Nature Rev. Genet.* **6**, 206–220 (2005).
- Kessler, A. & Baldwin, I. T. Plant responses to insect herbivory: the emerging molecular analysis. *Annu. Rev. Plant Biol.* **53**, 299–328 (2002).
- Gomez-Gomez, L. & Boller, T. Flagellin perception: a paradigm for innate immunity. *Trends Plant Sci.* **7**, 251–256 (2002).
- Felix, G., Duran, J. D., Volko, S. & Boller, T. Plants have a sensitive perception system for the most conserved domain of bacterial flagellin. *Plant J.* **18**, 265–276 (1999).
- Zipfel, C. *et al.* Bacterial disease resistance in *Arabidopsis* through flagellin perception. *Nature* **428**, 764–767 (2004).
- Chinchilla, D., Bauer, Z., Regenass, M., Boller, T. & Felix, G. The *Arabidopsis* receptor kinase FLS2 binds flg22 and determines the specificity of flagellin perception. *Plant Cell* **18**, 465–476 (2006).
- Robatzek, S., Chinchilla, D. & Boller, T. Ligand-induced endocytosis of the pattern recognition receptor FLS2 in *Arabidopsis*. *Genes Dev.* **20**, 537–542 (2006).
- Sun, W., Dunning, F. M., Pfund, C., Weingarten, R. & Bent, A. F. Within-species flagellin polymorphism in *Xanthomonas campestris* pv *campestris* and its impact on elicitation of *Arabidopsis* FLAGELLIN SENSING2-dependent defenses. *Plant Cell* **18**, 764–779, doi:10.1105/tpc.105.037648 (2006).
- Felix, G. & Boller, T. Molecular sensing of bacteria in plants. The highly conserved RNA-binding motif RNP-1 of bacterial cold shock proteins is recognized as an elicitor signal in tobacco. *J. Biol. Chem.* **278**, 6201–6208 (2003).
- Kunze, G. *et al.* The N terminus of bacterial elongation factor Tu elicits innate immunity in *Arabidopsis* plants. *Plant Cell* **16**, 3496–3507 (2004).
- Zipfel, C. *et al.* Perception of the bacterial PAMP EF-Tu by the receptor EFR restricts *Agrobacterium*-mediated transformation. *Cell* **125**, 749–760 (2006).
- Shiu, S. H. & Bleeker, A. B. Expansion of the receptor-like kinase/Pelle gene family and receptor-like proteins in *Arabidopsis*. *Plant Physiol.* **132**, 530–543 (2003).
- Dardick, C. & Ronald, P. Plant and animal pathogen recognition receptors signal through non-RD kinases. *PLoS Pathog* **2**, e2 (2006).
- Fritz-Laylin, L. K., Krishnamurthy, N., Tor, M., Sjolander, K. V. & Jones, J. D. Phylogenomic analysis of the receptor-like proteins of rice and *Arabidopsis*. *Plant Physiol.* **138**, 611–623 (2005).
- Schulze-Lefert, P. & Panstruga, R. Establishment of biotrophy by parasitic fungi and reprogramming of host cells for disease resistance. *Annu. Rev. Phytopathol.* **41**, 641–667 (2003).
- Badel, J. L., Charkowski, A. O., Deng, W. L. & Collmer, A. A gene in the *Pseudomonas syringae* pv. *tomato* Hrp pathogenicity island conserved effector locus, *hopPtoA1*, contributes to efficient formation of bacterial colonies in planta and is duplicated elsewhere in the genome. *Mol. Plant Microbe Interact.* **15**, 1014–1024 (2002).
- Grant, S. R., Fisher, E. J., Chang, J. H., Mole, B. M. & Dangl, J. L. Subterfuge and manipulation: type III effector proteins of phytopathogenic bacteria. *Annu. Rev. Microbiol.* **60**, 425–449, doi:10.1146/annurev.micro.60.080805.142251 (2006).
- Abramovitch, R. B., Anderson, J. C. & Martin, G. B. Bacterial elicitation and evasion of plant innate immunity. *Nature Rev. Mol. Cell Biol.* **7**, 601–611 (2006).
- Mudgett, M. B. New insights to the function of phytopathogenic bacterial type III effectors in plants. *Annu. Rev. Plant Biol.* **56**, 509–531 (2005).
- Jakobek, J. L., Smith, J. A. & Lindgren, P. B. Suppression of bean defense responses by *Pseudomonas syringae*. *Plant Cell* **5**, 57–63 (1993).
- Thilmony, R., Underwood, W. & He, S. Y. Genome-wide transcriptional analysis of the *Arabidopsis thaliana* interaction with the plant pathogen *Pseudomonas syringae* pv. *tomato* DC3000 and the human pathogen *Escherichia coli* O157:H7. *Plant J.* **46**, 34–53 (2006).
- Tao, Y. *et al.* Quantitative nature of *Arabidopsis* responses during compatible and incompatible interactions with the bacterial pathogen *Pseudomonas syringae*. *Plant Cell* **15**, 317–330 (2003).
- Truman, W., de Zabal, M. T. & Grant, M. Type III effectors orchestrate a complex interplay between transcriptional networks to modify basal defence responses during pathogenesis and resistance. *Plant J.* **46**, 14–33 (2006).
- Nomura, K., Melotto, M. & He, S. Y. Suppression of host defense in compatible plant–*Pseudomonas syringae* interactions. *Curr. Opin. Plant Biol.* **8**, 361–368 (2005).
- Desveaux, D., Singer, A. U. & Dangl, J. L. Type III effector proteins: doppelgangers of bacterial virulence. *Curr. Opin. Plant Biol.* **9**, 376–382 (2006).
- Nomura, K. *et al.* A bacterial virulence protein suppresses host innate immunity to cause plant disease. *Science* **313**, 220–223, doi:10.1126/science.1129523 (2006).
- DeRoy, S., Thilmony, R., Kwack, Y.-B., Nomura, K. & He, S. Y. A family of conserved bacterial effectors inhibits salicylic acid-mediated basal immunity and promotes disease necrosis in plants. *Proc. Natl Acad. Sci. USA* **101**, 9927–9932 (2004).
- He, P. *et al.* Specific bacterial suppressors of MAMP signaling upstream of MAPKKK in *Arabidopsis* innate immunity. *Cell* **125**, 563–575 (2006).
- Abramovitch, R. B., Kim, Y. J., Chen, S., Dickman, M. B. & Martin, G. B. *Pseudomonas* type III effector AvrPtoB induces plant disease susceptibility by inhibition of host programmed cell death. *EMBO J.* **22**, 60–69 (2003).
- de Torres, M. *et al.* *Pseudomonas syringae* effector AvrPtoB suppresses basal defence in *Arabidopsis*. *Plant J.* **47**, 368–382, 10.1111/j.1365-313X.2006.02798.x (2006).
- Janjusevic, R., Abramovitch, R. B., Martin, G. B. & Stebbins, C. E. A bacterial inhibitor of host programmed cell death defenses is an E3 ubiquitin ligase. *Science* **311**, 222–226 (2006).
- Mukherjee, S. *et al.* *Yersinia* YopJ acetylates and inhibits kinase activation by blocking phosphorylation. *Science* **312**, 1211–1214 (2006).
- Rivas, S. & Thomas, C. M. Molecular interactions between tomato and the leaf mold pathogen *Cladosporium fulvum*. *Annu. Rev. Phytopathol.* **43**, 395–436 (2005).
- Allen, R. L. *et al.* Host–parasite coevolutionary conflict between *Arabidopsis* and downy mildew. *Science* **306**, 1957–1960 (2004).
- Rehmany, A. P. *et al.* Differential recognition of highly divergent downy mildew avirulence gene alleles by *RPP1* resistance genes from two *Arabidopsis* lines. *Plant Cell* **17**, 1839–1850 (2005).
- Bhattacharjee, S. *et al.* The malarial host-targeting signal is conserved in the Irish potato famine pathogen. *PLoS Pathog* **2**, e50 (2006).
- Dodds, P. N., Lawrence, G. J., Catanzariti, A. M., Ayliffe, M. A. & Ellis, J. G. The *Melampsora lini* AvrL567 avirulence genes are expressed in haustoria and their products are recognized inside plant cells. *Plant Cell* **16**, 755–768 (2004).
- Catanzariti, A. M., Dodds, P. N., Lawrence, G. J., Ayliffe, M. A. & Ellis, J. G. Haustorially expressed secreted proteins from flax rust are highly enriched for avirulence elicitors. *Plant Cell* **18**, 243–256 (2006).
- Zhao, Y. *et al.* Virulence systems of *Pseudomonas syringae* pv. *tomato* promote bacterial speck disease in tomato by targeting the jasmonate signaling pathway. *Plant J.* **36**, 485–499 (2003).
- Brooks, D. M., Bender, C. L. & Kunkel, B. N. The *Pseudomonas syringae* phytotoxin coronatine promotes virulence by overcoming salicylic acid-dependent defences in *Arabidopsis thaliana*. *Mol. Plant Pathol.* **6**, 629–640 (2005).
- Melotto, M., Underwood, W., Koczan, J., Nomura, K. & He, S. The innate immune function of plant stomata against bacterial invasion. *Cell* **126**, 969–980 (2006).
- Navarro, L. *et al.* A plant miRNA contributes to antibacterial resistance by repressing auxin signaling. *Science* **312**, 436–439 (2006).
- Greenberg, J. T. & Yao, N. The role and regulation of programmed cell death in plant–pathogen interactions. *Cell. Microbiol.* **6**, 201–211 (2004).
- Schulze-Lefert, P. Plant immunity: the origami of receptor activation. *Curr. Biol.* **14**, R22–R24 (2004).

55. Holt, B. F. III, Belkhadir, Y. & Dangl, J. L. Antagonistic control of disease resistance protein stability in the plant immune system. *Science* **309**, 929–932 (2005).
56. Takken, F. L., Albrecht, M. & Tameling, W. I. Resistance proteins: molecular switches of plant defence. *Curr. Opin. Plant Biol.* **9**, 383–390 (2006).
57. Durrant, W. E. & Dong, X. Systemic acquired resistance. *Annu. Rev. Phytopathol.* **42**, 185–209 (2004).
58. Dorey, S. *et al.* Spatial and temporal induction of cell death, defense genes, and accumulation of salicylic acid in tobacco leaves reacting hypersensitively to a fungal glycoprotein. *Mol. Plant Microbe Interact.* **10**, 646–655 (1997).
59. Torres, M. A., Jones, J. D. & Dangl, J. L. Pathogen-induced, NADPH oxidase-derived reactive oxygen intermediates suppress spread of cell death in *Arabidopsis thaliana*. *Nature Genet.* **37**, 1130–1134 (2005).
60. Eulgem, T. Regulation of the *Arabidopsis* defense transcriptome. *Trends Plant Sci.* **10**, 71–78 (2005).
61. Kim, H. S. *et al.* The *Pseudomonas syringae* effector AvrRpt2 cleaves its C-terminally acylated target, RIN4, from *Arabidopsis* membranes to block RPM1 activation. *Proc. Natl Acad. Sci. USA* **102**, 6496–6501 (2005).
62. Mackey, D., Holt, B. F. III, Wiig, A. & Dangl, J. L. RIN4 interacts with *Pseudomonas syringae* type III effector molecules and is required for RPM1-mediated disease resistance in *Arabidopsis*. *Cell* **108**, 743–754 (2002).
63. Axtell, M. J., Chisholm, S. T., Dahlbeck, D. & Staskawicz, B. J. Genetic and molecular evidence that the *Pseudomonas syringae* type III effector protein AvrRpt2 is a cysteine protease. *Mol. Microbiol.* **49**, 1537–1546 (2003).
64. Coaker, G., Falick, A. & Staskawicz, B. Activation of a phytopathogenic bacterial effector protein by a eukaryotic cyclophilin. *Science* **308**, 548–550 (2005).
65. Chisholm, S. T. *et al.* Molecular characterization of proteolytic cleavage sites of the *Pseudomonas syringae* effector AvrRpt2. *Proc. Natl Acad. Sci. USA* **102**, 2087–2092 (2005).
66. Mackey, D., Belkhadir, Y., Alonso, J. M., Ecker, J. R. & Dangl, J. L. *Arabidopsis* RIN4 is a target of the type III virulence effector AvrRpt2 and modulates RPS2-mediated resistance. *Cell* **112**, 379–389 (2003).
67. Axtell, M. J. & Staskawicz, B. J. Initiation of RPS2-specified disease resistance in *Arabidopsis* is coupled to the AvrRpt2-directed elimination of RIN4. *Cell* **112**, 369–377 (2003).
68. Day, B., Dahlbeck, D. & Staskawicz, B. J. NDR1 interaction with RIN4 mediates the differential activation of multiple disease resistance pathways. *Plant Cell* (September 29, doi:10.1105/tpc.106.044693 2006).
69. Belkhadir, Y., Nimchuk, Z., Hubert, D. A., Mackey, D. & Dangl, J. L. *Arabidopsis* RIN4 negatively regulates disease resistance mediated by RPS2 and RPM1 downstream or independent of the NDR1 signal modulator, and is not required for the virulence functions of bacterial type III effectors AvrRpt2 or AvrRpm1. *Plant Cell* **16**, 2822–2835 (2004).
70. Day, B. *et al.* Molecular basis for the RIN4 negative regulation of RPS2 disease resistance. *Plant Cell* **17**, 1292–1305 (2005).
72. Dodds, P. N. *et al.* Direct protein interaction underlies gene-for-gene specificity and coevolution of the flax resistance genes and flax rust avirulence genes. *Proc. Natl Acad. Sci. USA* **103**, 8888–8893 (2006).
73. Jia, Y., McAdams, S. A., Bryan, G. T., Hershey, H. P. & Valent, B. Direct interaction of resistance gene and avirulence gene products confers rice blast resistance. *EMBO J.* **19**, 4004–4014 (2000).
74. Deslandes, L. *et al.* Physical interaction between RRS1-R, a protein conferring resistance to bacterial wilt, and PopP2, a type III effector targeted to the plant nucleus. *Proc. Natl Acad. Sci. USA* **100**, 8024–8029 (June 3, doi:10.1073/pnas.1230660100 2003).
75. Lipka, V. *et al.* Pre- and postinvasion defenses both contribute to nonhost resistance in *Arabidopsis*. *Science* **310**, 1180–1183, doi:10.1126/science.1119409 (2005).
76. Stein, M. *et al.* *Arabidopsis* PEN3/PDR8, an ATP binding cassette transporter, contributes to nonhost resistance to inappropriate pathogens that enter by direct penetration. *Plant Cell* **18**, 731–746, doi:10.1105/tpc.105.038372 (2006).
77. Yun, B. W. *et al.* Loss of actin cytoskeletal function and EDS1 activity, in combination, severely compromises non-host resistance in *Arabidopsis* against wheat powdery mildew. *Plant J.* **34**, 768–777 (2003).
78. Feys, B. J. *et al.* *Arabidopsis* SENESCENCE-ASSOCIATED GENE101 stabilizes and signals within an ENHANCED DISEASE SUSCEPTIBILITY1 complex in plant innate immunity. *Plant Cell* **17**, 2601–2613 (2005).
79. Collins, N. C. *et al.* SNARE-protein-mediated disease resistance at the plant cell wall. *Nature* **425**, 973–977 (2003).
80. Assaad, F. F. *et al.* The PEN1 syntaxin defines a novel cellular compartment upon fungal attack and is required for the timely assembly of papillae. *Mol. Biol. Cell* **15**, 5118–5129 (2004).
81. Bhat, R. A., Miklis, M., Schmelzer, E., Schulze-Lefert, P. & Panstruga, R. Recruitment and interaction dynamics of plant penetration resistance components in a plasma membrane microdomain. *Proc. Natl Acad. Sci. USA* **102**, 3135–3140 (2005).
82. Consonni, C. *et al.* Conserved requirement for a plant host cell protein in powdery mildew pathogenesis. *Nature Genet.* **38**, 716–720 (2006).
83. Kobayashi, D. Y., Tamaki, S. J. & Keen, N. T. Cloned avirulence genes from the tomato pathogen *Pseudomonas syringae* pv. *tomato* confer cultivar specificity on soybean. *Proc. Natl Acad. Sci. USA* **86**, 157–161 (1989).
84. Lorang, J. M., Shen, H., Kobayashi, D., Cooksey, D. & Keen, N. T. *avrA* and *avrE* in *Pseudomonas syringae* pv. *tomato* PT23 play a role in virulence on tomato plants. *Mol. Plant Microbe Interact.* **7**, 208–215 (1994).
85. Peyyala, R. & Farman, M. L. *Magnaporthe oryzae* isolates causing gray leaf spot of perennial ryegrass possess a functional copy of the *AVR1-CO39* avirulence gene. *Mol. Plant Pathol.* **7**, 157–165 (2006).
86. Staal, J., Kaliff, M., Bohman, S. & Dixelius, C. Transgressive segregation reveals two *Arabidopsis* TIR-NB-LRR resistance genes effective against *Leptosphaeria maculans*, causal agent of blackleg disease. *Plant J.* **46**, 218–230 (2006).
87. Rohmer, L., Guttman, D. S. & Dangl, J. L. Diverse evolutionary mechanisms shape the type III effector virulence factor repertoire in the plant pathogen *Pseudomonas syringae*. *Genetics* **167**, 1341–1360 (2004).
88. Tsiamis, G. *et al.* Cultivar-specific avirulence and virulence functions assigned to *avrPphF* in *Pseudomonas syringae* pv. *phaseolicola*, the cause of bean halo-blight disease. *EMBO J.* **19**, 3204–3214 (2000).
89. Lawrence, G. L., Mayo, G. M. E. & Shepherd, K. W. Interactions between genes controlling pathogenicity in the flax rust fungus. *Phytopathol.* **71**, 12–19 (1981).
90. Bakker, E. G., Toomajian, C., Kreitman, M. & Bergelson, J. A. Genome-wide survey of *R* gene polymorphisms in *Arabidopsis*. *Plant Cell* **18**, 1803–1818 (2006).
91. Kuang, H., Woo, S. S., Meyers, B. C., Nevo, E. & Michelmore, R. W. Multiple genetic processes result in heterogeneous rates of evolution within the major cluster disease resistance genes in lettuce. *Plant Cell* **16**, 2870–2894 (2004).
92. Van der Hoorn, R. A., De Wit, P. J. & Joosten, M. H. Balancing selection favors guarding resistance proteins. *Trends Plant Sci.* **7**, 67–71 (2002).
93. Thrall, P. H. & Burdon, J. J. Evolution of virulence in a plant host-pathogen metapopulation. *Science* **299**, 1735–1737 (2003).
94. Tian, D., Traw, M. B., Chen, J. Q., Kreitman, M. & Bergelson, J. Fitness costs of *R*-gene-mediated resistance in *Arabidopsis thaliana*. *Nature* **423**, 74–77 (2003).
95. Kemen, E. *et al.* Identification of a protein from rust fungi transferred from haustoria into infected plant cells. *Mol. Plant Microbe Interact.* **18**, 1130–1139 (2005).
96. Szurek, B., Marois, E., Bonas, U. & Van den Ackerveken, G. Eukaryotic features of the *Xanthomonas* type III effector AvrBs3: protein domains involved in transcriptional activation and the interaction with nuclear importin receptors from pepper. *Plant J.* **26**, 523–534 (2001).
97. Holt, B. F. III. *et al.* An evolutionarily conserved mediator of plant disease resistance gene function is required for normal *Arabidopsis* development. *Dev. Cell* **2**, 807–817 (2002).
98. Palma, K., Zhang, Y. & Li, X. An importin α homolog, MOS6, plays an important role in plant innate immunity. *Curr. Biol.* **15**, 1129–1135 (2005).
99. Zhang, Y. & Li, X. A putative nucleoporin 96 is required for both basal defense and constitutive resistance responses mediated by *suppressor of npr1-1, constitutive 1*. *Plant Cell* **17**, 1306–1316, doi: 10.1105/tpc.104.029926 (2005).
100. Shao, F. *et al.* Cleavage of *Arabidopsis* PBS1 by a bacterial type III effector. *Science* **301**, 1230–1233 (2003).
101. Mucyn, T. S. *et al.* The NB-ARC-LRR protein Prf interacts with Pto kinase *in vivo* to regulate specific plant immunity. *Plant Cell* (October 6, doi:10.1105/tpc.106.044016 2006).
102. Abramovitch, R. B. & Martin, G. B. AvrPtoB: a bacterial type III effector that both elicits and suppresses programmed cell death associated with plant immunity. *FEMS Microbiol. Lett.* **245**, 1–8 (2005).
103. Wu, A. J., Andriotis, V. M., Durrant, M. C. & Rathjen, J. P. A patch of surface-exposed residues mediates negative regulation of immune signaling by tomato Pto kinase. *Plant Cell* **16**, 2809–2821 (2004).
104. Rooney, H. C. *et al.* Cladosporium Avr2 inhibits tomato Rcr3 protease required for Cf-2-dependent disease resistance. *Science* **308**, 1783–1786 (2005).

Acknowledgements We thank S. Grant, D. Hubert, J. McDowell, M. Nishimura, R. Panstruga, J. Rathjen, P. Schulze-Lefert, S. Somerville and C. Zipfel for their critical comments, and our various colleagues for permission to cite their unpublished results. We apologize to the many colleagues whose work could not be discussed owing to strict page limits. Our labs are supported by grants from the NIH, NSF and DOE (JLD) and from the Gatsby Charitable Trust, the BBSRC and the EU (JDGJ).

Author Contributions JDGJ and JLD conceived and wrote the review together.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence should be addressed to J.J. (jonathan.jones@tsl.ac.uk) or J.L.D. (dangl@email.unc.edu).

ARTICLES

Analysis of one million base pairs of Neanderthal DNA

Richard E. Green¹, Johannes Krause¹, Susan E. Ptak¹, Adrian W. Briggs¹, Michael T. Ronan², Jan F. Simons², Lei Du², Michael Egholm², Jonathan M. Rothberg², Maja Paunovic^{3,†} & Svante Pääbo¹

Neanderthals are the extinct hominid group most closely related to contemporary humans, so their genome offers a unique opportunity to identify genetic changes specific to anatomically fully modern humans. We have identified a 38,000-year-old Neanderthal fossil that is exceptionally free of contamination from modern human DNA. Direct high-throughput sequencing of a DNA extract from this fossil has thus far yielded over one million base pairs of hominoid nuclear DNA sequences. Comparison with the human and chimpanzee genomes reveals that modern human and Neanderthal DNA sequences diverged on average about 500,000 years ago. Existing technology and fossil resources are now sufficient to initiate a Neanderthal genome-sequencing effort.

Neanderthals were first recognized as a distinct group of hominids from fossil remains discovered 150 years ago at Feldhofer in Neander Valley, outside Düsseldorf, Germany. Subsequent Neanderthal finds in Europe and western Asia showed that fossils with Neanderthal traits appear in the fossil record of Europe and western Asia about 400,000 years ago and vanish about 30,000 years ago. Over this period they evolved morphological traits that made them progressively more distinct from the ancestors of modern humans that were evolving in Africa^{1,2}. For example, the crania of late Neanderthals have protruding mid-faces, brain cases that bulge outward at the sides, and features of the base of the skull, jaw and inner ears that set them apart from modern humans³.

The nature of the interaction between Neanderthals and modern humans, who expanded out of Africa around 40,000–50,000 years ago and eventually replaced Neanderthals as well as other archaic hominids across the Old World is still a matter of some debate. Although there is no evidence of contemporaneous cohabitation at any single site, there is evidence of geographical and temporal overlap in their ranges before the disappearance of Neanderthals. Additionally, late in their history, some Neanderthal groups adopted cultural traits such as body decorations, potentially through cultural interactions with incoming modern humans⁴.

In 1997, a segment of the hypervariable control region of the maternally inherited mitochondrial DNA (mtDNA) of the Neanderthal type specimen found at Feldhofer was sequenced. Phylogenetic analysis showed that it falls outside the variation of contemporary humans and shares a common ancestor with mtDNAs of present-day humans approximately half a million years ago^{5,6}. Subsequently, mtDNA sequences have been retrieved from eleven additional Neanderthal specimens: Feldhofer 2 in Germany⁷, Mezmaiskaya in Russia⁸, Vindija 75, 77 and 80 in Croatia^{9,10}, Engis 2 in Belgium, La Chapelle-aux-Saints and Rochers de Villeneuve in France¹⁰, Scladina in Belgium¹¹, Monte Lessini in Italy¹², and El Sidron 441 in Spain¹³. Although some of these sequences are extremely short, they are all more closely related to one another than to modern human mtDNAs^{9,11}.

This fact, in conjunction with the absence of any related mtDNA sequences in currently living humans or in a small number of early modern human fossils^{5,10} strongly suggests that Neanderthals contributed no

mtDNA to present-day humans. On the basis of various population models, it has been estimated that a maximal overall genetic contribution of Neanderthals to the contemporary human gene pool is between 25% and 0.1% (refs 10, 14). Because the latter conclusions are based on mtDNA, a single maternally inherited locus, they are limited in their ability to detect a Neanderthal contribution to the current human gene pool both by the vagaries of genetic drift and by the possibility of a sex bias in reproduction. However, both morphological evidence^{4,15} and the variation in the modern human gene pool¹⁶ support the conclusion that if any genetic contribution of Neanderthals to modern human occurred, it was of limited magnitude.

Neanderthals are the hominid group most closely related to currently living humans, so a Neanderthal nuclear genome sequence would be an invaluable resource for annotating the human genome. Roughly 35 million nucleotide differences exist between the genomes of humans and chimpanzees, our closest living relatives¹⁷. Soon, genome sequences from other primates such as the orang-utan and the macaque will allow such differences to be assigned to the human and chimpanzee lineages. However, temporal resolution of the genetic changes along the human lineage, where remarkable morphological, behavioural and cognitive changes occurred, are limited without a more closely related genome sequence for comparison. In particular, comparison to the Neanderthal would enable the identification of genetic changes that occurred during the last few hundred thousand years, when fully anatomically and behaviourally modern humans appeared.

Identification of a Neanderthal fossil for DNA sequencing

Although it is possible to recover mtDNA¹⁸ and occasionally even nuclear DNA sequences^{19–22} from well-preserved remains of organisms that are less than a few hundred thousand years old, determination of ancient hominid sequences is fraught with special difficulties and pitfalls¹⁸. In addition to degradation and chemical damage to the DNA that can cause any ancient DNA to be irretrievable or misread, contamination of specimens, laboratory reagents and instruments with traces of DNA from modern humans must be avoided. In fact, when sensitive polymerase chain reaction (PCR) is used, human

¹Max-Planck Institute for Evolutionary Anthropology, Deutscher Platz 6, D-04103 Leipzig, Germany. ²454 Life Sciences, 20 Commercial Street, Branford, Connecticut 06405, USA.

³Institute of Quaternary Paleontology and Geology, Croatian Academy of Sciences and Arts, A. Kovacica 5/II, HR-10 000 Zagreb, Croatia.

[†]Deceased.

mtDNA sequences can be retrieved from almost every ancient specimen^{23,24}. This problem is especially severe when Neanderthal remains are studied because Neanderthal and human are so closely related that one expects to find few or no differences between Neanderthals and modern humans within many regions²⁵, making it impossible to rely on the sequence information itself to distinguish endogenous from contaminating DNA sequences. A necessary first step for sequencing nuclear DNA from Neanderthals is therefore to identify a Neanderthal specimen that is free or almost free of modern human DNA.

We tested more than 70 Neanderthal bone and tooth samples from different sites in Europe and western Asia for bio-molecular preservation by removing samples of a few milligrams for amino acid analysis. The vast majority of these samples had low overall contents of amino acids and/or high levels of amino acid racemization, a stereoisomeric structural change that affects amino acids in fossils, indicating that they are unlikely to contain retrievable endogenous DNA²⁶. However, some of the samples are better preserved in that they contain high levels of amino acids (more than 20,000 p.p.m.), low levels of racemization of amino acids such as aspartate that racemize rapidly, as well as amino acid compositions that suggest that the majority of the preserved protein stems from collagen.

From 100–200 mg of bone from six of these specimens we extracted DNA and analysed the relative abundance of Neanderthal-like mtDNA sequences and modern human-like mtDNA sequences by performing PCR with primer pairs that amplify both human and Neanderthal mtDNA with equal efficiency. The amplification products span segments of the hypervariable region of the mtDNA in which all Neanderthals sequenced to date differ from all contemporary humans. From subsequent cloning into a plasmid vector and sequencing of more than a hundred clones from each product, we determined the ratio of Neanderthal-like to modern human-like mtDNA in each extract. We used two different primer pairs that amplify fragments of 63 base pairs and 119 base pair to gauge the contamination levels for different lengths of DNA molecules.

Figure 1 shows that the level of contamination differs drastically among the samples. Whereas only around 1% of the mtDNA present in three samples from France, Russia and Uzbekistan was Neanderthal-like, one sample from Croatia and one from Spain contained around 5% and 75% Neanderthal-like mtDNA, respectively. One bone (Vi-80) from Vindija Cave, Croatia, stood out in that ~99% of the 63-base-pair mtDNA segments and ~94% of the 119-base pair segments are of Neanderthal origin. Assuming that the ratio of Neanderthal to contaminating modern human DNA is the same for mtDNA as it is for nuclear DNA, the Vi-80 bone therefore yields DNA fragments that are predominantly of Neanderthal

origin and provided that the contamination rate was not increased during the downstream sequencing process, the extent of contamination in the final analyses is below ~6%.

The Vi-80 bone was discovered by M. Malez and co-workers in layer G3 of Vindija Cave in 1980. It has been dated by carbon-14 accelerator mass spectrometry to $38,310 \pm 2,130$ years before present and its entire mtDNA hypervariable region I has been sequenced¹⁰. Out of 14 Neanderthal remains from layer G3 that we have analysed, this bone is one of six samples that show good bio-molecular preservation, while the other eight bones show intermediate to bad states of preservation that do not suggest the presence of amplifiable DNA. Preservation conditions in Vindija Cave thus vary drastically from bone to bone, a situation that may be due to different extents of water percolation in different parts of the cave.

Direct large-scale DNA sequencing from the Vindija Neanderthal

Because the Vi-80 Neanderthal bone extract is largely free of contaminating modern human mtDNA, we chose this extract to perform large-scale parallel 454 sequencing²⁷. In this technology, single-stranded libraries, flanked by common adapters, are created from the DNA sample and individual library molecules are amplified through bead-based emulsion PCR, resulting in beads carrying millions of clonal copies of the DNA fragments from the samples. These are subsequently sequenced by pyrosequencing on the GS20 454 sequencing system.

For several reasons, the 454 sequencing platform is extremely well suited for analyses of bulk DNA extracted from ancient remains²⁸. First, it circumvents bacterial cloning, in which the vast majority of initial template molecules are lost during transformation and establishment of clones. Second, because each molecule is amplified in isolation from other molecules it also precludes template competition, which frequently occurs when large numbers of different DNA fragments are amplified together. Third, its current read length of 100–200 nucleotides covers the average length of the DNA preserved in most fossils²⁹. Fourth, it generates hundreds of thousands of reads per run, which is crucial because the majority of the DNA recovered from fossils is generally not derived from the fossil species, but rather from organisms that have colonized the organism after its death^{20,30}. Fifth, because each sequenced product stems from just one original single-stranded template molecule of known orientation, the DNA strand from which the sequence is derived is known²⁸. This provides an advantage over traditional PCR from double-stranded templates, in which the template strand is not known, because the frequency of different nucleotide misincorporations can be deduced. For example, using 454 sequencing, the rate at which cytosine is converted to uracil and read as thymine can be distinguished from the rate at which guanine is converted to xanthine and read as adenine, whereas this is impossible using traditional PCR or bacterial cloning. This is important since nucleotide conversions and misincorporations in ancient DNA are caused by damage that affects different bases differently^{28,31} and this pattern of false substitutions can be used to estimate the relative probability that a particular substitution (that is, the observation of a nucleotide difference between DNA sequences) represents the authentic DNA sequence of the organism versus an artefact from DNA degradation.

We recovered a total of 254,933 unique sequences from the Vi-80 bone (see Supplementary Methods). These were aligned to the human (build 36.1)³², chimpanzee (build 1)¹⁷ and mouse (build 34.1)³³ complete genome sequences, to environmental sample sequences in the GenBank *env* database (version 3, September 2005), and to the complete set of redundant nucleotide sequences in GenBank *nt* (version 3, September 2005, excluding EST, STS, GSS, environmental and HTGS sequences)³⁴ using the program BLASTN (NCBI version 2.2.12)³⁵. The most similar database sequence for each query was identified and classified by its taxonomic order (Fig. 2) (see Supplementary Methods). No significant nucleotide sequence similarity in the databases was found for 79% of the fossil extract

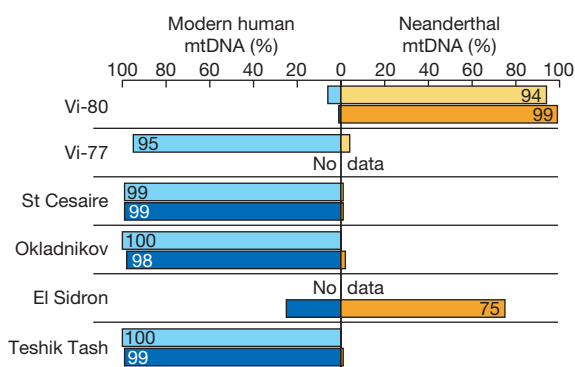


Figure 1 | Ratio of Neanderthal to modern human mtDNA in six hominid fossils. For each fossil, primer pairs that amplify a long (119 base pairs; upper lighter bars) and short (63 base pairs; lower darker bars) product were used to amplify segments of the mtDNA hypervariable region. The products were sequenced and determined to be either of Neanderthal (yellow) or modern human (blue) type.

sequence reads. This is typical of large-scale sequencing both from other ancient bones^{20,22,28} and from environmental samples^{36,37}, although some permafrost-preserved specimens can yield high amounts of endogenous DNA²². Sequences with similarity to a database sequence were classified by the taxonomic order of their most significant alignment. Actinomycetales, a bacterial order with many soil-living species, was the most populous order and accounted for 6.8% of the sequences. The second most populous order, to which 15,701 unique sequences or 6.2% of the sequence reads were most similar, was that of primates. All other individual orders were substantially less frequent. Notably, the average percentage identity for the primate sequence alignments was 98.8%, whereas it was 92–98% for the other frequently occurring orders. Thus, the primate reads, unlike many of the prokaryotic reads, are aligned to a very closely related species.

Neanderthal mtDNA sequences

Among the 15,701 sequences of primate origin, we first identified all mtDNA in order to investigate whether their evolutionary relationship to the current human mtDNA pool is similar to what is known from previous analyses of Neanderthal mtDNA. A total of 41 unique DNA sequences from the Vi-80 fossil had their closest hits to different parts of the human mtDNA, and comprised, in total, 2,705 base pairs of unique mtDNA sequence. None of the putative Neanderthal mtDNA sequences map to the two hypervariable regions that have been previously sequenced in Neanderthals. We aligned these mtDNA sequences to the complete mtDNA sequences of 311 modern humans from different populations³⁸ as well as to the complete mtDNA sequences of three chimpanzees and two bonobos (Supplementary Information). A schematic neighbour-joining tree estimated from this alignment is shown in Fig. 3. In agreement with previous results, the Neanderthal mtDNA falls outside the variation among modern humans. However, the length of the branch leading to the Neanderthal mtDNA is 2.5 times as long as the branch leading to modern human mtDNAs. This is likely to be due to errors in our Neanderthal sequences derived from substitution artefacts from damaged, ancient DNA and from sequencing errors²⁸.

To analyse the extent to which errors occur in the Neanderthal mtDNA reads, we designed 29 primer pairs (Supplementary Methods) flanking all 39 positions at which the Vi-80 Neanderthal mtDNA sequences differed by substitutions from the consensus bases seen among the 311 human mtDNA sequences. These primer pairs,

which are designed to yield amplification products that vary in length between 50 and 98 base pairs (including primers), were used in a multiplex two-step PCR³⁹ from the same Neanderthal extract that had been used for large-scale 454 sequencing. Twenty five of the PCR products, containing 34 of the positions where the Neanderthal differs from humans, were successfully amplified and cloned, and then six or more clones of each product were sequenced. The consensus sequence seen among these clones revealed the same nucleotides seen by the 454 sequencing at 20 of the 34 positions and no additional differences. Of the 14 positions found to represent errors in the sequence reads, seven were C to T transitions, four were G to A, two were G to T and one was T to C. This pattern of change is typical for ancient DNA, where deamination of cytosine residues³¹ and, to a lesser extent, modifications of guanosine residues²⁸ have been found to account for the majority of nucleotide misincorporations during PCR.

These results also show that the likelihood of observing errors in the sequencing reads is drastically different depending on whether one considers nucleotide positions where a base in the Neanderthal mtDNA sequence differs from both the human and chimpanzee sequences, or positions where the Neanderthal differ from the humans but is identical to the chimpanzee mtDNA sequences. Among the mtDNA sequences analysed, there are 14 positions where the Neanderthal carries a base identical to the chimpanzee, and 13 of those were confirmed by PCR. In contrast, among the remaining 20 positions, where the Neanderthal sequences differed from both humans and chimpanzees, only seven were confirmed. When only PCR-confirmed sequence data are used to estimate the mtDNA tree (Fig. 3), the Neanderthal branch has a length comparable to that of contemporary humans. This suggests that no large source of errors other than what is detected by the PCR analysis affects the sequences.

Using these PCR-confirmed substitutions and a divergence time between humans and chimpanzees of 4.7–8.4 million years^{40–42}, we estimate the divergence time for the mtDNA fragments determined here to be 461,000–825,000 years. This is in general agreement with previous estimates of Neanderthal–human mtDNA divergence of 317,000–741,000 years⁶ based on mtDNA hypervariable region sequences and is compatible with our presumption that the mtDNA sequences determined from the Vi-80 extract are of Neanderthal origin.

Nuclear DNA sequences

We next analysed the sequence reads whose closest matches are to the human or chimpanzee nuclear genomes and that are at least 30 base

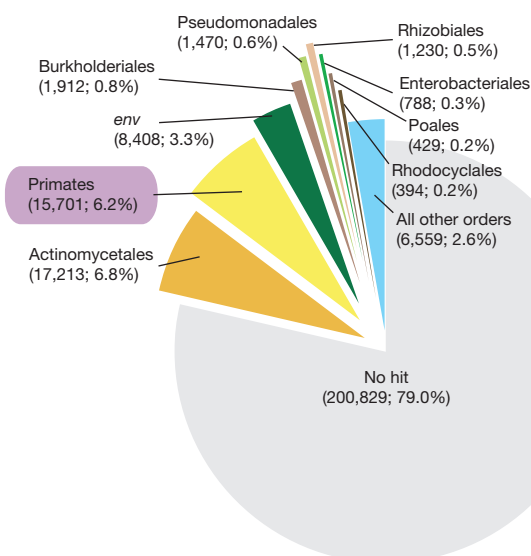


Figure 2 | Taxonomic distribution of DNA sequences from the Vi-80 extract. The taxonomic order of the database sequence giving the best alignment for each unique sequence read was determined. The most populous taxonomic orders are shown.

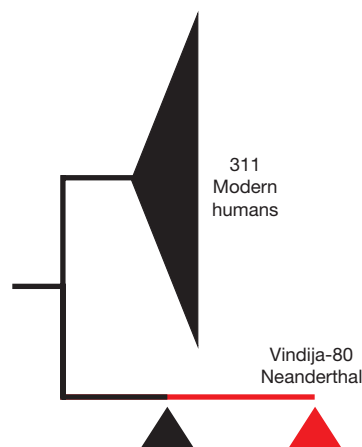


Figure 3 | Schematic tree relating the Vi-80 Neanderthal mtDNA sequences to 311 human mtDNA sequences. The Neanderthal branch length is given with uncorrected sequences (red triangle) and after correction of sequences via independent PCRs (black triangle). Chimpanzee and bonobo sequences (not shown) were used to root the neighbour-joining tree. Several substitution models (Kimura 2-parameter, Tajima-Nei, and Tamura 3-parameter with uniform or gamma-distributed ($\gamma = 0.5$ –1.1) rates) yielded bootstrap support values for the human branch from 72–83%.

pairs long. Figure 4 shows where they map to the human karyotype (see Supplementary Methods). Overall, 0.04% of the autosomal genome sequence is covered by the Neanderthal reads—on average 3.61 bases per 10,000 bases. Both X and Y chromosomes are represented, with a lower coverage of 2.18 and 1.62 bases per 10,000, respectively, showing that the Vi-80 bone is derived from a male individual.

The data presented in Fig. 4 show that when the hit density for sequences that have a single best hit in the human genome is plotted along the chromosomes, several suggestive local deviations from the average hit density are seen, which may represent copy-number differences in the Neanderthal relative to the human reference genome. For comparison, we generated 454 sequence data from a DNA sample from a modern human. Interestingly, some of the deviations seen in the Neanderthal are present also in the modern human, whereas others are not. The latter group of sequences may indicate copy-number differences that are unique to the Neanderthal relative to the modern human genome sequence. Thus, when more Neanderthal sequence is generated in the future, it may be possible to determine copy number differences between the Neanderthal, the chimpanzee and the human genomes.

Patterns of nucleotide change on lineages

We generated three-way alignments between all Neanderthal sequences that map uniquely within the human genome and the corresponding human and chimpanzee genome sequences (see

Supplementary Methods). An important artefact of local sequence alignments, such as those produced here, is that they necessarily begin and end with regions of exact sequence identity. The size of these regions is a function of the scoring parameters for the alignment. In this case, five bases at both ends of the alignments, amounting to ~14% of all data, needed to be removed (Supplementary Fig. 1) to eliminate biases in estimates of sequence divergence.

Each autosomal nucleotide position in the alignment that did not contain a deletion in the Neanderthal, the human or the chimpanzee sequences and was associated with a chimpanzee genome position with quality score ≥ 30 was classified according to which species share the same bases (Fig. 5). A total of 736,941 positions contained the same base in all three groups. The next largest category comprises 10,167 positions in which the human and Neanderthal base are identical, but the chimpanzee base is different. These positions are likely to have changed either on the hominid lineage before the divergence between human and Neanderthal sequences or on the chimpanzee lineage. At 3,447 positions, the Neanderthal base differs from both the human and chimpanzee bases, which are identical to each other. As suggested by the analysis of the mtDNA sequences, this category contains positions that have changed on the Neanderthal lineage, as well as a large proportion of errors that derive both from base damage that have accumulated in the ancient DNA and from sequencing errors. At 434 positions, the human base differs from both the Neanderthal and chimpanzee bases, which are identical to each other.

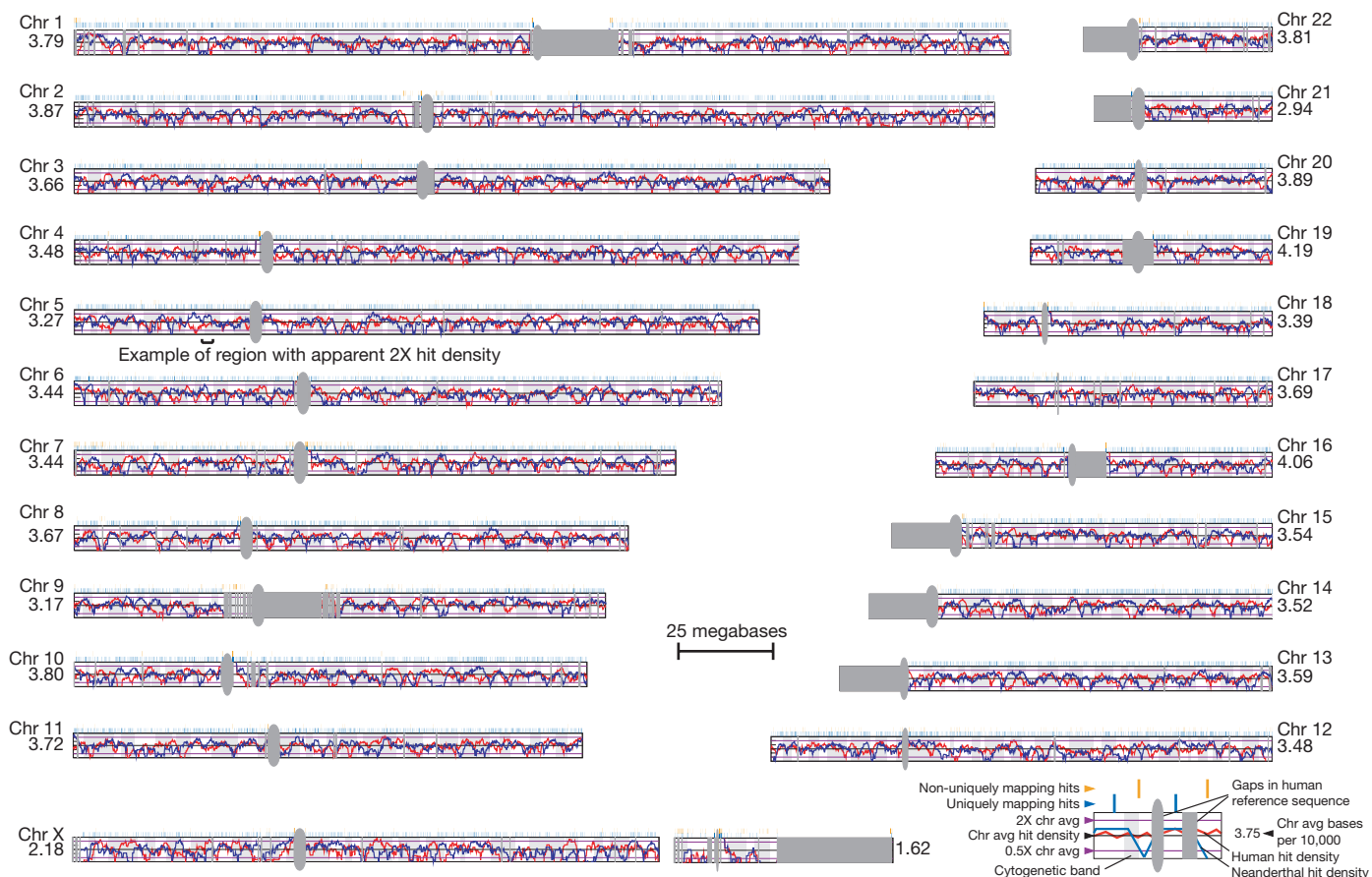


Figure 4 | Location on the human karyotype of Neanderthal DNA sequences. All sequences longer than 30 nucleotides whose best alignments were to the human genome are shown. The blue lines above each chromosome mark the position of all alignments that are unique in terms of bit-score within the human genome. Orange lines are alignments that have more than one alignment of equal bit-score. To the left of each chromosome, the average number of Neanderthal bases per 10,000 is given. Lines (Neanderthal, blue; human, red) within each chromosome show the hit

density, on a log-base 2 scale, within sliding windows of 3 megabases along each chromosome. The centre black lines indicate the average hit-density for the chromosomes. The purple lines above and below indicate hit densities of 2X and 1/2X the chromosome average, respectively. On chromosome 5, an example of a region of increased sequence density is highlighted. Sequence gaps in the human reference sequence are indicated by dark grey regions. Chromosomal banding pattern is indicated by light grey regions.

These positions are likely to have changed on the human lineage after the divergence from Neanderthal. Finally, a total of 51 positions contain different bases in all three groups.

Because the 454 sequencing technology allows the base in a base pair from which a sequence is derived to be determined, the relative frequencies of each of the 12 possible categories of base changes can be estimated for each evolutionary lineage. As seen in Fig. 5, the patterns of the chimpanzee-specific and human-specific changes are similar to each other in that the eight transversional changes are of approximately equal frequency and about fourfold less frequent than each of the four transitional changes, yielding a transition to transversion ratio of 2.04, typical of closely related mammalian genomes⁴³. For the Neanderthal-specific changes the pattern is very different in that mismatches are dominated by C to T and G to A differences. Thus, the pattern of change seen among the Neanderthal-specific alignment mismatches is typical of the nucleotide substitution pattern observed in PCR of ancient DNA.

Consistent with this, modern human sequences determined by 454 sequencing show no excess amount of C to T or G to A differences (Supplementary Fig. 2), indicating that lesions in the ancient DNA rather than sequencing errors account for the majority of the errors in the Neanderthal sequences. Assuming that the evolutionary rate of DNA change was the same on the Neanderthal and human lineages, the majority of observed differences specific to the Neanderthal lineage are artefacts. All Neanderthal-specific changes were therefore disregarded in the subsequent analyses and the Neanderthal sequences were used solely to assign changes to the human or chimpanzee lineage where the human and chimpanzee genome sequences differ and the Neanderthal sequence carries either the human or the chimpanzee base.

Genomic divergence between Neanderthals and humans

Assuming that the rates of DNA sequence change along the chimpanzee lineage and the human lineage were similar, it can be estimated that 8.2% of the DNA sequence changes that have occurred on the human lineage since the divergence from the chimpanzee lineage occurred after the divergence of the Neanderthal lineage. However, although the Neanderthal-specific changes that are heavily influenced by errors are not used for this analysis, some errors in the single-pass sequencing reads from the Neanderthal extract will create positions where the Neanderthal is identical either to human or chimpanzee sequences, and thus affect the estimates of sequence change on the human and chimpanzee lineages. When the effects

of such errors in the Neanderthal sequences are quantified and removed (see Supplementary Methods), ~7.9% of the sequence changes along the human lineage are estimated to have occurred after divergence from the Neanderthal. If the human–chimpanzee divergence time is set to 6,500,000 years (refs 40, 41, 44), this implies an average human–Neanderthal DNA sequence divergence time of ~516,000 years. A 95% confidence interval generated by bootstrap re-sampling of the alignment data gives a range of 465,000 to 569,000 years. Obviously, these divergence estimates are dependent on the human–chimpanzee divergence time, which is a much larger source of uncertainty.

We analysed the DNA sequences generated from a contemporary human using the same sequencing protocol as was used for the Neanderthal. Although ancient DNA is degraded and damaged, this comparison controls for many of the aspects of the analysis including sequencing and alignment methodology. In this case, ~7.1% of the divergence along the human lineage is assigned to the time subsequent to the divergence of the two human sequences. The average divergence time between alleles within humans is thus ~459,000 years with a 95% confidence interval between 419,000 and 498,000 years. As expected, this estimate of the average human diversity is less than the divergence seen between the human and the Neanderthal sequences, but constitutes a large fraction of it because much of the human sequence diversity is expected to predate the human–Neanderthal split²⁵. Neanderthal genetic differences to humans must therefore be interpreted within the context of human diversity.

Ancestral population size

Humans differ from apes in that their effective population size is of the order of 10,000 while those of chimpanzees, gorillas and orangutans are two to four times larger^{45–47}. Furthermore, the population size of the ancestor of humans and chimpanzees was found to be similar to those of apes, rather than to humans^{42,48}. The Neanderthal sequence data now allow us to ask if the effective size of the population ancestral to humans and Neanderthals was large, as is the case for apes and the human–chimpanzee ancestor, or small, as for present-day humans.

We applied a method⁴² that co-estimates the ancestral effective population size and the split time between Neanderthal and human populations (Fig. 6a; see Supplementary Methods). As seen in Fig. 6b, we recover a line describing combinations of population sizes and split times compatible with the data and lack power to be more

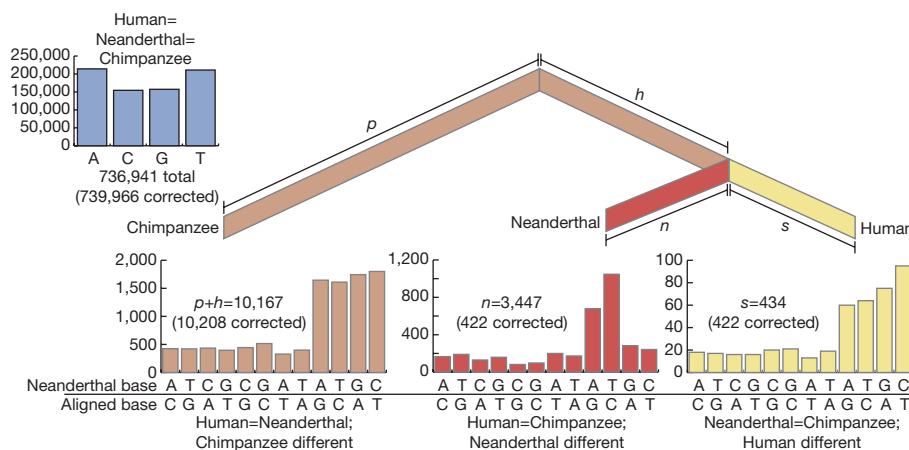


Figure 5 | Schematic tree illustrating the number of nucleotide changes inferred to have occurred on hominoid lineages. In blue is the distribution of all aligned positions that did not change on any lineage. In brown are the changes that occurred either on the chimpanzee lineage (p) or on the hominid lineage (h) before the human and Neanderthal lineages diverged. In red are the changes that are unique to the Neanderthal lineage (n), including

all changes due to base-damage and base-calling errors. In yellow are changes unique to the human lineage. The distributions of types of changes in each category are also given. The numbers of changes in each category, corrected for base-calling errors in the Neanderthal sequence (see Supplementary Methods), are shown within parentheses.

precise (see Supplementary Methods and Results). Using this line we can estimate the ancestral population size, given estimates about the population split time from independent sources. If we use a split time of 400,000 years inferred from the fossil record (J. J. Hublin, personal communication), then our point estimate of the ancestral population size is $\sim 3,000$. Given uncertainty in both the sequence divergence time and the population split time, our estimate of the ancestral population size varies from 0 to 12,000.

These results suggest that the population ancestral to present-day humans and Neanderthals was similar to present-day humans in having a small effective size and thus that the effective population size on the hominid lineage had already decreased before the split between humans and Neanderthals. Therefore, the small effective population size seen in present-day human samples may not be unique to modern humans, but was present also in the common ancestor of Neanderthals and modern humans. We speculate that a small effective size, perhaps associated with numerous expansions from small groups, was typical not only of modern humans but of many groups of the genus *Homo*. In fact, the origin of *Homo erectus* may have been associated with genetic or cultural adaptations that resulted in drastic population expansions as indicated by their appearance outside Africa around two million years ago.

Neanderthal sequences and human polymorphisms

Another question that can be addressed with these data is how often the Neanderthal has the ancestral allele (that is, the same allele seen in the chimpanzee) versus the derived (or novel) allele at sites where

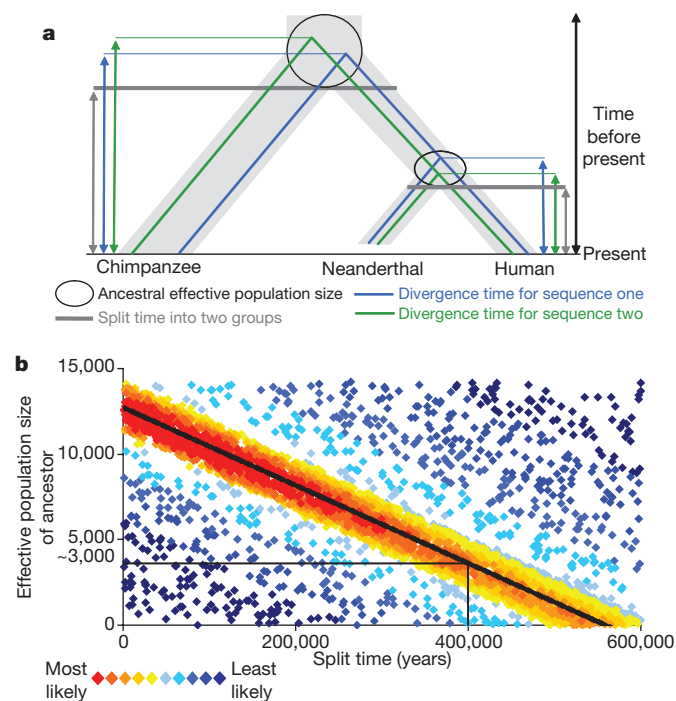


Figure 6 | Estimate of the effective population size of the ancestor of humans and Neanderthals. **a**, Schematic illustration of the model used to estimate ancestral effective population size. By split time, we mean the time, in the past, after which there was no more interbreeding between two groups. By divergence, we mean the time, in the past, at which two genetic regions separated and began to accumulate substitutions independently. Effective population size is the number of individuals needed under ideal conditions to produce the amount of observed genetic diversity within a population. **b**, The likelihood estimates of population split times and ancestral population sizes. The likelihoods are grouped by colour. The red–yellow points are statistically equivalent based on the likelihood ratio test approximation. The black line is the line of best fit to red–yellow points (see Supplementary Methods). This graph is scaled assuming a human–chimpanzee average sequence divergence time of 6,500,000 years.

humans carry a single nucleotide polymorphism (SNP). The latter case identifies SNPs that were present in the common ancestor of Neanderthals and present-day humans. Using the SNPs that overlap with our data from two large genome-wide data sets (HapMap⁴⁹, 786 SNPs and Perlegen⁵⁰, 318 SNPs), we find that the Neanderthal sample has the derived allele in $\sim 30\%$ of all SNPs. This number is presumably an overestimate since the SNPs analysed were ascertained to be of high frequency in present-day humans and hence are more likely to be old. Nevertheless, this high level of derived alleles in the Neanderthal is incompatible with the simple population split model estimated in the previous section, given split times inferred from the fossil record. This may suggest gene flow between modern humans and Neanderthals. Given that the Neanderthal X chromosome shows a higher level of divergence than the autosomes (R.E.G., unpublished observation), gene flow may have occurred predominantly from modern human males into Neanderthals. More extensive sequencing of the Neanderthal genome is necessary to address this possibility.

Rationale and prospects for a Neanderthal genome sequence

We demonstrate here that DNA sequences can be generated from the Neanderthal nuclear genome by massive parallel sequencing on the 454 sequencing platform. It is thus feasible to determine large amounts of sequences from this extinct hominid. As a corollary, it is possible to envision the determination of a Neanderthal genome sequence. For several reasons, we believe that this would represent a valuable genomic resource.

First, a Neanderthal genome sequence would allow all nucleotide sequence differences as well as many copy-number differences between the human and chimpanzee genomes to be temporally resolved with respect to whether they occurred before the separation of humans from Neanderthals, or whether they occurred after or at the time of separation. The latter class of changes is of interest, because some of them will be associated with the emergence of modern humans. A Neanderthal genome sequence would therefore allow the research community to determine whether DNA sequence differences between humans and chimpanzees that are found to be functionally important represent recent changes on the human lineage. No data other than a Neanderthal genome sequence can provide this information.

Second, the fact that Neanderthals carry the derived allele for a substantial fraction of human SNPs suggests a method of identifying genomic regions that have experienced a selective sweep subsequent to the separation of human and Neanderthal populations. Such selective sweeps in the human genome will make the variation in these regions younger than the separation of humans and Neanderthals. As we show above, in regions not affected by sweeps a substantial proportion of polymorphic sites in humans will carry derived alleles in the Neanderthal genome sequence, whereas no sites will do so in regions affected by sweeps. This represents an approach to identifying selective sweeps in humans that is not possible from other data.

Third, once large amounts of Neanderthal genome sequence is generated, it will become possible to estimate the misincorporation probabilities for each class of nucleotide differences between the Neanderthal and chimpanzee genomes with high accuracy by analysing regions covered by many reads such as mtDNA, repeated genome regions of high sequence identity, as well as single-copy regions covered by multiple reads. Once this is done, the confidence that any particular nucleotide position where the Neanderthal differs from human as well as chimpanzee is correct can be reliably estimated. In combination with future knowledge about the function of genes and biological systems, comprehensive information from the Neanderthal genome will then allow aspects of Neanderthal biology to be deciphered that are unavailable by any other means.

Are fossil and technical resources today sufficient to imagine the determination of a Neanderthal genome sequence? The results presented here are derived from approximately one fifteenth of an extract

prepared from ~100 mg of bone. To achieve one-fold coverage of the Neanderthal genome (3 gigabases) without any further improvement in technology, about twenty grams of bone and 6,000 runs on the current version of the 454 sequencing platform would be necessary. Although this is at present a daunting task, technical improvements in the procedures described here that would make the retrieval of DNA sequences of the order of ten times more efficient can easily be envisioned (our unpublished results). In view of that prospect, we have recently initiated a project that aims at achieving an initial draft version of the Neanderthal genome within two years.

Received 14 July; accepted 11 October 2006.

1. Bischoff, J. L. *et al.* The Sima de los Huesos hominids date to beyond U/Th equilibrium (>350 kyr) and perhaps to 400–500 kyr: New radiometric dates. *J. Archaeol. Sci.* **30**, 275–280 (2003).
2. Hublin, J.-J. (ed.) *Climatic Changes, Paleogeography, and the Evolution of the Neandertals* (Plenum Press, New York, 1998).
3. Franciscus, R. G. (ed.) *Neanderthals* (Oxford Univ. Press, Oxford, 2002).
4. Hublin, J. J., Spoor, F., Braun, M., Zonneveld, F. & Condemi, S. A late Neanderthal associated with Upper Palaeolithic artefacts. *Nature* **381**, 224–226 (1996).
5. Krings, M. *et al.* Neanderthal DNA sequences and the origin of modern humans. *Cell* **90**, 19–30 (1997).
6. Krings, M., Geisert, H., Schmitz, R. W., Krainitzki, H. & Pääbo, S. DNA sequence of the mitochondrial hypervariable region II from the Neanderthal type specimen. *Proc. Natl Acad. Sci. USA* **96**, 5581–5585 (1999).
7. Schmitz, R. W. *et al.* The Neanderthal type site revisited: Interdisciplinary investigations of skeletal remains from the Neander Valley, Germany. *Proc. Natl Acad. Sci. USA* **99**, 13342–13347 (2002).
8. Ovchinnikov, I. V. *et al.* Molecular analysis of Neanderthal DNA from the northern Caucasus. *Nature* **404**, 490–493 (2000).
9. Krings, M. *et al.* A view of Neanderthal genetic diversity. *Nature Genet.* **26**, 144–146 (2000).
10. Serre, D. *et al.* No evidence of Neanderthal mtDNA contribution to early modern humans. *PLoS Biol.* **2**, 313–317 (2004).
11. Orlando, L. *et al.* Revisiting Neanderthal diversity with a 100,000 year old mtDNA sequence. *Curr. Biol.* **16**, R400–R402 (2006).
12. Caramelli, D. *et al.* A highly divergent mtDNA sequence in a Neanderthal individual from Italy. *Curr. Biol.* **16**, R630–R632 (2006).
13. Lalueza-Fox, C. *et al.* Neanderthal evolutionary genetics: mitochondrial DNA data from the Iberian peninsula. *Mol. Biol. Evol.* **22**, 1077–1081 (2005).
14. Currat, M. & Excoffier, L. Modern humans did not admix with Neanderthals during their range expansion into Europe. *PLoS Biol.* **2**, e421 (2004).
15. Stringer, C. Modern human origins: progress and prospects. *Phil. Trans. R. Soc. Lond. B* **357**, 563–579 (2002).
16. Takahata, N., Lee, S. H. & Satta, Y. Testing multiregionality of modern human origins. *Mol. Biol. Evol.* **18**, 172–183 (2001).
17. Chimpanzee Sequencing and Analysis Consortium. Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature* **437**, 69–87 (2005).
18. Pääbo, S. *et al.* Genetic analyses from ancient DNA. *Annu. Rev. Genet.* **38**, 645–679 (2004).
19. Greenwood, A. D., Capelli, C., Possnert, G. & Paabo, S. Nuclear DNA sequences from late Pleistocene megafauna. *Mol. Biol. Evol.* **16**, 1466–1473 (1999).
20. Noonan, J. P. *et al.* Genomic sequencing of Pleistocene cave bears. *Science* **309**, 597–599 (2005).
21. Rompler, H. *et al.* Nuclear gene indicates coat-color polymorphism in mammoths. *Science* **313**, 62 (2006).
22. Poinar, H. N. *et al.* Metagenomics to paleogenomics: large-scale sequencing of mammoth DNA. *Science* **311**, 392–394 (2006).
23. Hofreiter, M., Serre, D., Poinar, H. N., Kuch, M. & Pääbo, S. Ancient DNA. *Nature Rev. Genet.* **2**, 353–359 (2001).
24. Malmstrom, H., Stora, J., Dalen, L., Holmlund, G. & Gotherstrom, A. Extensive human DNA contamination in extracts from ancient dog bones and teeth. *Mol. Biol. Evol.* **22**, 2040–2047 (2005).
25. Pääbo, S. Human evolution. *Trends Cell Biol.* **9**, M13–M16 (1999).
26. Poinar, H. N., Höss, M., Bada, J. L. & Pääbo, S. Amino acid racemization and the preservation of ancient DNA. *Science* **272**, 864–866 (1996).
27. Margulies, M. *et al.* Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437**, 376–380 (2005).
28. Stiller, M. *et al.* Patterns of nucleotide misincorporations during enzymatic amplification and direct large-scale sequencing of ancient DNA. *Proc. Natl Acad. Sci. USA* **103**, 13578–13584 (2006).
29. Pääbo, S. Ancient DNA: extraction, characterization, molecular cloning, and enzymatic amplification. *Proc. Natl Acad. Sci. USA* **86**, 1939–1943 (1989).
30. Höss, M., Dilling, A., Currant, A. & Pääbo, S. Molecular phylogeny of the extinct ground sloth *Myiodon darwini*. *Proc. Natl Acad. Sci. USA* **93**, 181–185 (1996).
31. Hofreiter, M., Jaenicke, V., Serre, D., Haeseler, A. & Pääbo, S. DNA sequences from multiple amplifications reveal artifacts induced by cytosine deamination in ancient DNA. *Nucleic Acids Res.* **29**, 4793–4799 (2001).
32. International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).
33. Mouse Genome Sequencing Consortium. Initial sequencing and comparative analysis of the mouse genome. *Nature* **420**, 520–562 (2002).
34. Benson, D. A., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J. & Wheeler, D. L. GenBank. *Nucleic Acids Res.* **34**, D16–D20 (2006).
35. Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).
36. Beja, O. *et al.* Construction and analysis of bacterial artificial chromosome libraries from a marine microbial assemblage. *Environ. Microbiol.* **2**, 516–529 (2000).
37. Venter, J. C. *et al.* Environmental genome shotgun sequencing of the Sargasso Sea. *Science* **304**, 66–74 (2004).
38. Ingman, M. & Gyllenstein, U. mtDB: Human Mitochondrial Genome Database, a resource for population genetics and medical sciences. *Nucleic Acids Res.* **34**, D749–D751 (2006).
39. Krause, J. *et al.* Multiplex amplification of the mammoth mitochondrial genome and the evolution of Elephantidae. *Nature* **439**, 724–727 (2006).
40. Kumar, S., Filipski, A., Swarna, V., Walker, A. & Blair Hedges, S. Placing confidence limits on the molecular age of the human-chimpanzee divergence. *Proc. Natl Acad. Sci. USA* **102**, 18842–18847 (2005).
41. Patterson, N., Richter, D. J., Gnerre, S., Lander, E. S. & Reich, D. Genetic evidence for complex speciation of humans and chimpanzees. *Nature* **441**, 1103–1108 (2006).
42. Wall, J. D. Estimating ancestral population sizes and divergence times. *Genetics* **163**, 395–404 (2003).
43. Yang, Z. & Yoder, A. D. Estimation of the transition/transversion rate bias and species sampling. *J. Mol. Evol.* **48**, 274–283 (1999).
44. Innan, H. & Watanabe, H. The effect of gene flow on the coalescent time in the human-chimpanzee ancestral population. *Mol. Biol. Evol.* **23**, 1040–1047 (2006).
45. Kaessmann, H., Wiebe, V., Weiss, G. & Pääbo, S. Great ape DNA sequences reveal a reduced diversity and an expansion in humans. *Nature Genet.* **27**, 155–156 (2001).
46. Yu, N., Jensen-Seaman, M. I., Chernick, L., Ryder, O. & Li, W.-H. Nucleotide diversity in gorillas. *Genetics* **166**, 1375–1383 (2004).
47. Fischer, A., Pollack, J., Thalmann, O., Nickel, B. & Pääbo, S. Demographic history and genetic differentiation in apes. *Curr. Biol.* **16**, 1133–1138 (2006).
48. Rannala, B. & Yang, Z. Bayes estimation of species divergence times and ancestral population sizes using DNA sequences from multiple loci. *Genetics* **164**, 1645–1656 (2003).
49. The International HapMap Consortium. A haplotype map of the human genome. *Nature* **437**, 1299–1320 (2005).
50. Hinds, D. A. *et al.* Whole genome patterns of common DNA variation in diverse human populations. *Science* **307**, 1072–1079 (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We are indebted to G. Coop, W. Enard, I. Hellmann, A. Fischer, P. Johnson, S. Kudaravalli, M. Lachmann, T. Maricic, J. Pritchard, J. Noonan, D. Reich, E. Rubin, M. Slatkin, L. Vigilant and T. Weaver for discussions. We thank A. P. Derevianko, C. Lalueza-Fox, A. Rosas and B. Vandermeersch for fossil samples. We also thank the Croatian Academy of Sciences and Arts for support and the Innovation Fund of Max Planck Society for financial support. 454 Life Sciences thanks NHGRI for continued support for the development of this platform, as well as all of its employees who developed the sequencing system. R.E.G. is supported by an NSF postdoctoral fellowship in Biological Informatics.

Author Contributions M.P. provided Neanderthal samples and palaeontological information; J.M.R. and S.P. conceived of and initiated the 454 Neanderthal sequencing approach; M.T.R. developed the library preparation method, and generated and processed the sequencing data; J.F.S. planned and coordinated library preparation and sequencing activities; L.D. processed and transferred data between 454 Life Sciences and the MPI; M.E. supervised, planned and coordinated research between MPI and 454 Life Sciences; J.K. and A.W.B. extracted ancient DNA and performed analyses in the “Identification of a Neanderthal fossil for DNA sequencing” section; J.K. and R.E.G. performed analyses in the “Neanderthal mtDNA sequences” section; R.E.G. performed the analyses in the sections “Direct large-scale DNA sequencing” to “Genomic divergence between Neanderthals and humans”; S.E.P. performed analyses in the sections “Ancestral population size” and “Neanderthal sequences and human polymorphisms”; S.P. conceived of the ideas presented in the section “Rationale and prospects for a Neanderthal genome sequence”, and initiated, planned and coordinated the study; R.E.G., S.E.P., J.K. and S.P. wrote the paper.

Author Information Neanderthal fossil extract sequences were deposited at EBI with accession numbers CAAN01000001–CAAN01369630. Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details accompany the paper on www.nature.com/nature. Correspondence and requests for materials should be addressed to R.E.G. (green@eva.mpg.de).

Resveratrol improves health and survival of mice on a high-calorie diet

Joseph A. Baur^{1*}, Kevin J. Pearson^{2*}, Nathan L. Price², Hamish A. Jamieson⁷, Carles Lerin⁸, Avash Kalra², Vinayakumar V. Prabhu³, Joanne S. Allard², Guillermo Lopez-Lluch⁹, Kaitlyn Lewis², Paul J. Pistell², Suresh Poosala⁴, Kevin G. Becker³, Olivier Boss¹⁰, Dana Gwinn¹¹, Mingyi Wang⁵, Sharan Ramaswamy⁶, Kenneth W. Fishbein⁶, Richard G. Spencer⁶, Edward G. Lakatta⁵, David Le Couteur⁷, Reuben J. Shaw¹¹, Placido Navas⁹, Pere Puigserver⁸, Donald K. Ingram^{2,12}, Rafael de Cabo² & David A. Sinclair¹

Resveratrol (3,5,4'-trihydroxystilbene) extends the lifespan of diverse species including *Saccharomyces cerevisiae*, *Caenorhabditis elegans* and *Drosophila melanogaster*. In these organisms, lifespan extension is dependent on Sir2, a conserved deacetylase proposed to underlie the beneficial effects of caloric restriction. Here we show that resveratrol shifts the physiology of middle-aged mice on a high-calorie diet towards that of mice on a standard diet and significantly increases their survival. Resveratrol produces changes associated with longer lifespan, including increased insulin sensitivity, reduced insulin-like growth factor-1 (IGF-I) levels, increased AMP-activated protein kinase (AMPK) and peroxisome proliferator-activated receptor- γ coactivator 1 α (PGC-1 α) activity, increased mitochondrial number, and improved motor function. Parametric analysis of gene set enrichment revealed that resveratrol opposed the effects of the high-calorie diet in 144 out of 153 significantly altered pathways. These data show that improving general health in mammals using small molecules is an attainable goal, and point to new approaches for treating obesity-related disorders and diseases of ageing.

The number of overweight individuals worldwide has reached 2.1 billion, leading to an explosion of obesity-related health problems associated with increased morbidity and mortality^{1,2}. Although the association of obesity with increased risk of cardiovascular disease and diabetes is well known, it is often under-appreciated that the risks of other age-related diseases, such as cancer and inflammatory disorders, are also increased. At the other end of the spectrum, reducing caloric intake by ~40% below that of *ad libitum*-fed animals (caloric restriction) is the most robust and reproducible way to delay age-related diseases and extend lifespan in mammals^{3,4}.

Experiments with *Saccharomyces cerevisiae* and *Drosophila melanogaster* have implicated the sirtuin/Sir2 family of NAD⁺-dependent deacetylases and mono-ADP-ribosyltransferases as mediators of the physiological effects of caloric restriction⁵. In mammals, seven sirtuin genes have been identified (*SIRT1*–*7*). *SIRT1* regulates such processes as glucose and insulin production, fat metabolism, and cell survival, leading to speculation that sirtuins might mediate effects of caloric restriction in mammals⁵. We previously screened over 20,000 molecules to identify ~25 that enhance *SIRT1* activity *in vitro*⁶. Resveratrol, a molecule produced by a variety of plants in response to stress, emerged as the most potent.

Resveratrol has since been shown to extend the lifespan of evolutionarily distant species including *S. cerevisiae*, *C. elegans* and *D. melanogaster* in a Sir2-dependent manner^{6–9}. A recent study found that resveratrol improves health and extends maximum lifespan by 59% in a vertebrate fish¹⁰. In mammalian cells, resveratrol produces

SIRT1-dependent effects that are consistent with improved cellular function and organismal health^{11–15}. Whether resveratrol acts directly or indirectly through Sir2 *in vivo* is currently a subject of debate¹⁶.

On the basis of the unprecedented ability of resveratrol to improve health and extend lifespan in simple organisms, we have asked whether it has similar effects in mice. We hypothesized that resveratrol might shift the physiology of mice on a high-calorie diet towards that of mice on a standard diet and provide the associated health benefits without the mice having to reduce caloric intake. Cohorts of middle-aged (one-year-old) male C57BL/6NIA mice were provided with either a standard diet (SD) or an otherwise equivalent high-calorie diet (60% of calories from fat, HC) for the remainder of their lives. To each of the diets, we added resveratrol at two concentrations that provided an average of 5.2 ± 0.1 and 22.4 ± 0.4 mg kg⁻¹ day⁻¹, which are feasible daily doses for humans. After 6 months of treatment, there was a clear trend towards increased survival and insulin sensitivity. Because the effects were more prominent in the higher dose (22.4 ± 0.4 mg kg⁻¹ day⁻¹, HCR), we initially focused our resources on this group and present the results of those analyses herein. Analyses of the other groups will be presented at a later date.

Increased survival

Mice on the HC diet steadily gained weight until ~75 weeks of age, after which average weight slowly declined (Fig. 1a). Although mice on the HCR diet were slightly lighter than the HC mice during the initial months, there was no significant weight difference between the

¹Department of Pathology, Paul F. Glenn Laboratories for the Biological Mechanisms of Aging, Harvard Medical School, 77 Avenue Louis Pasteur, Boston, Massachusetts 02115, USA.

²Laboratory of Experimental Gerontology, ³Gene Expression and Genomics Unit, ⁴Research Resources Branch, ⁵Laboratory of Cardiovascular Science, and ⁶Laboratory of Clinical Investigation, Research Resources Branch of the Gerontology Research Center, National Institute on Aging, National Institutes of Health, 5600 Nathan Shock Drive, Baltimore, Maryland 21224, USA. ⁷Centre for Education and Research on Ageing, and the ANZAC Research Institute University of Sydney, Concord, New South Wales 2139, Australia.

⁸Department of Cell Biology, Johns Hopkins University School of Medicine, Baltimore, Maryland 21205, USA. ⁹Centro Andaluz de Biología del Desarrollo, Universidad Pablo de Olavide-CSIC, 41013 Sevilla, Spain. ¹⁰Sirtis Pharmaceuticals, Inc., 790 Memorial Drive, Cambridge, Massachusetts 02139, USA. ¹¹Molecular and Cell Biology Laboratory, The Salk Institute, 10010 N Torrey Pines Road, La Jolla, California 92037, USA. ¹²Nutritional Neuroscience and Aging Laboratory, Pennington Biomedical Research Center, Louisiana State University System, 6400 Perkins Road, Baton Rouge, Louisiana 70808, USA.

*These authors contributed equally to this work.

groups from 18–24 months, when most of our analyses were performed. There was also no difference in body temperature (Table 1), food consumption (Supplementary Fig. 1a, b), total faecal output or lipid content (Supplementary Fig. 1c, d), or post-mortem body fat distribution (Supplementary Fig. 2).

At 60 weeks of age, the survival curves of the HC and HCR groups began to diverge and have remained separated by a 3–4-month interval (Fig. 1b). A similar effect on survival was observed in a previous study of one-year-old C57BL/6 mice on caloric restriction, ultimately resulting in a 20% extension of mean lifespan¹⁷. With the present age of the colony at 114 weeks, 58% of the HC control animals have died (median lifespan 108 weeks), as compared to 42% of the HCR group and 42% of the SD controls. Although we cannot yet confidently predict the ultimate mean lifespan extension, Cox proportional hazards regression shows that resveratrol reduced the risk of death from the HC diet by 31% (hazard ratio = 0.69, $P = 0.020$), to a point where it was not significantly different from the SD group (hazard ratio = 1.03, $P = 0.88$).

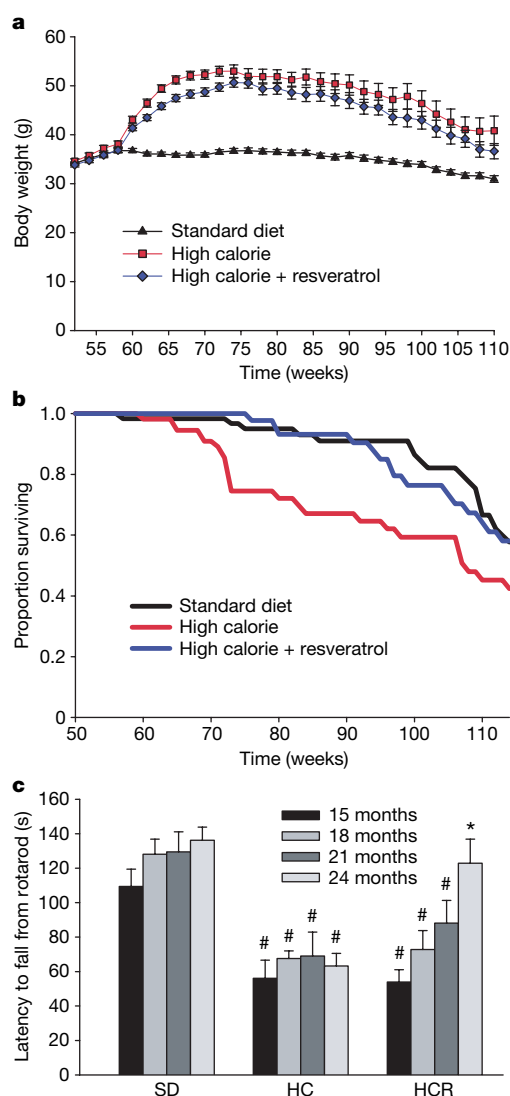


Figure 1 | Resveratrol increases survival and improves rotarod performance. **a**, Body weights of mice fed a standard diet (SD), high-calorie diet (HC), or high-calorie diet plus resveratrol (HCR). **b**, Kaplan–Meier survival curves. Hazard ratio for HCR is 0.69 ($\chi^2 = 5.39$, $P = 0.020$) versus HC, and 1.03 ($\chi^2 = 0.022$, $P = 0.88$) versus SD. The hazard ratio for HC versus SD is 1.43 ($\chi^2 = 5.75$, $P = 0.016$). **c**, Time to fall from an accelerating rotarod was measured every 3 months for all survivors from a pre-designated subset of each group; $n = 15$ (SD), 6 (HC) and 9 (HCR). Asterisk, $P < 0.05$ versus HC; hash, $P < 0.05$ versus SD. Error bars indicate s.e.m.

Although resveratrol increased survival, it was important to ascertain whether quality of life was maintained. One way to assess this was to measure balance and motor coordination, which we did by examining the ability to perform on a rotarod. Surprisingly, the resveratrol-fed HC mice steadily improved their motor skills as they aged, to the point where they were indistinguishable from the SD group (Fig. 1c). It is possible that the improved rotarod performance might have been due to minor differences in body weight but we view this as unlikely because we found no correlation between body weight and performance within groups and rotarod performance was also improved for resveratrol-treated SD mice (R.deC. and K.P., unpublished data). These data are reminiscent of the resveratrol-mediated increase in motor activity in older individuals of the vertebrate fish species *Nothobranchius furzeri*¹⁰.

Increased insulin sensitivity

In humans, high-calorie diets cause numerous pathological conditions including increased glucose and insulin levels leading to diabetes, cardiovascular disease and non-alcoholic fatty liver disease, a condition for which there is no effective treatment¹⁸. The HC-fed mice had alterations in plasma levels of markers that predict the onset of diabetes and a shorter lifespan, including increased levels of insulin, glucose and IGF-1 (Table 1). The HCR group had significantly lower levels of these markers, paralleling the SD group. An oral glucose tolerance test indicated that the insulin sensitivity of the resveratrol-treated mice was considerably higher than controls (Fig. 2a–d). Homeostatic model assessment, which is used to quantify insulin resistance, gave scores of 2.5 for SD, 8.8 for HC and 3.5 for HCR, confirming improved sensitivity (HCR versus HC, $P = 0.01$). Although the persistence of high glucose levels for more than 60 min following an oral dose is unusual for young mice, it is typical for older animals¹⁹. Compared to the HC controls, the areas under the curves for both glucose and insulin levels were significantly decreased in the resveratrol-fed HC group and were not significantly different from mice in the SD group (Fig. 2b, d).

Next, we investigated possible mechanisms behind these metabolic effects. AMPK is a metabolic regulator that promotes insulin sensitivity and fatty acid oxidation. Its activity correlates tightly with phosphorylation at Thr 172 (p-AMPK). Chronic activation of AMPK occurs on a calorically restriction diet and has been proposed as a longevity strategy for mammals²⁰. Consistent with this idea, additional copies of the AMPK gene are sufficient to extend lifespan in *C. elegans*²¹. Because we and others²² have observed that resveratrol can activate AMPK in cultured cells through an indirect mechanism (Fig. 2e; see also Supplementary 3a–d), we examined whether AMPK activation occurred in the livers of the resveratrol-fed group. Resveratrol showed a strong tendency towards inducing phosphorylation of AMPK (Fig. 2f), as well as two downstream indicators of activity, namely phosphorylation of acetyl-coA carboxylase at Ser 79 and decreased expression of fatty acid synthase (Supplementary Fig. 3e, f).

Decreased organ pathology

At 18 months of age it was apparent that the high-calorie diet greatly increased the size and weight of livers and that resveratrol prevented these changes (Fig. 3a–c; see also Supplementary Fig. 4a, b) without altering plasma lipid levels (Table 1). Histological examination of liver sections by staining with haematoxylin and eosin or oil red O revealed a loss of cellular integrity and the accumulation of large lipid droplets in the livers of the HC but not the HCR group. Blinded scoring of the liver sections for overall pathology on a scale of 0–4 (with 4 being the most severe) gave mean values of 1.3 for the SD group, 2.8 for the HC group and 0.8 for the HCR group (Fig. 3b). Plasma amylase, which can indicate pancreatic damage, was elevated in the HC group and was significantly reduced by resveratrol (Table 1). The reasons for the elevation of plasma amylase levels in the HC group are unclear given that pancreatic sections of all animals

revealed no damage to the pancreas or decrease in islet area (data not shown). Differences in the weights of other organs did not reach statistical significance.

The ability of resveratrol to improve motor function and increase insulin sensitivity indicated that its effects were not confined to the liver. To test directly whether other organs also benefited, we examined heart tissue of the SD, HC and HCR mice. Blinded scoring of overall pathology—taking into account subtle changes in the abundance of fatty lesions, cardiac muscle vacuolization, degeneration and inflammation—on a relative scale of 0–4 (with 4 being the most severe) gave mean values of 1.6 for the SD group, 3.2 for the HC group and 1.2 for HCR group (Fig. 3d; see also Supplementary Fig. 4c). Improvements in the morphology of the aortic elastic lamina were also apparent (Supplementary Fig. 4d).

Increased mitochondria

Exercise and reduced caloric intake increase hepatic mitochondrial number^{23,24} and we wondered whether resveratrol might produce the same effect. The livers of the resveratrol-treated mice had considerably more mitochondria than those of HC controls and were not significantly different compared to those of the SD group (Fig. 3e, f). There was also a trend towards higher citrate synthase activity in the resveratrol-fed mice (an indicator of increased mitochondrial content) although the effect was not significant (Table 1). Culturing FaO hepatoma or HeLa cells in the presence of resveratrol increased mitochondrial number (Fig. 3g, h), similar to the previously reported effect of culturing cells in serum from calorically restricted rats²⁴.

Mitochondrial biogenesis in liver and muscle is controlled, in large part, by the transcriptional coactivator PGC-1 α ^{25,26}, the activity of which, in turn, is positively regulated by SIRT1-mediated deacetylation^{27,28}. Hence, the acetylation status of PGC-1 α is considered a marker of SIRT1 activity *in vivo*²⁷. Because this assay required more tissue than was available, we examined a separate cohort of one-year-old mice on the HC diet that had been treated with resveratrol for 6 weeks at 186 mg kg⁻¹ d⁻¹. The acetylation status of PGC-1 α in the resveratrol-fed mice was threefold lower than the diet-matched controls (Fig. 3i, j). There was no detectable increase in SIRT1 protein

levels in resveratrol-treated mice (data not shown), suggesting that SIRT1 enzymatic activity was enhanced by resveratrol.

Microarrays and pathway analysis

These data demonstrate that resveratrol can alleviate the negative impact of a high-calorie diet on overall health and lifespan. To determine to what extent resveratrol had shifted the physiology of the high-calorie group towards the lower calorie group, we performed whole-genome microarrays and pathway analysis on liver samples. Z ratios were calculated as described previously²⁹ and a subset of expression changes was verified by polymerase chain reaction with reverse transcription (RT-PCR) (Supplementary Fig. 5). In the HCR group, expression patterns for 782 out of 41,534 (<2%) individual genes changed significantly relative to the diet-matched controls (Fig. 4a, b). Notably, within the top 12 most highly elevated transcripts were serum amyloid proteins (*Saa1–3*), major urinary proteins (*Mup1* and *Mup3*), and both forms of hydroxysteroid dehydrogenase that degrade testosterone (*Hsd3b4*, *Hsd3b5*). The list of 12 most highly downregulated transcripts included three cytochrome p450 enzymes (*Cyp2a4*, *Cyp2a5* and *Cyp2b9*) that are known to activate pro-carcinogens³⁰. The complete data set is available at <http://www.grc.nia.nih.gov/branches/rrb/dna/index/dnapubs.htm#2>.

We next performed parametric analysis of gene set enrichment (PAGE), a computational method that determines differences between pathways using *a priori* defined gene sets^{31,32}. PAGE analysis indicated that resveratrol caused a significant alteration in 127 pathways, including the TCA cycle, glycolysis, the classic and alternative complement pathways, butanoate and propanoate metabolism, sterol biosynthesis and Stat3 signalling (Supplementary Fig. 6; for a complete list see Supplementary Fig. 7). Some of the most highly downregulated pathways in the resveratrol-fed group are known to extend lifespan in model organisms when attenuated, including insulin signalling, IGF-1 and mTOR signalling, oxidative phosphorylation and electron transport^{33–36}. Downregulation of glycolysis is a well known marker of caloric restriction³⁷ and has been proposed as a mechanism by which caloric restriction works³⁸. The increase in Stat3, a transcription factor involved in cell survival and liver

Table 1 | Effects of a high-fat diet and resveratrol on various biomarkers in plasma

Parameter	Standard diet	High calorie	High calorie + resveratrol	Fed or Fasted
Free fatty acids (mequiv.)	0.27 (0.04) 0.83 (0.10)	0.59 (0.06)† 0.45 (0.20)	0.53 (0.03)† 0.54 (0.05)	Fed Fasted
Triglycerides (mg dl ⁻¹)	76.6 (6.8)	81.4 (6.6)	88.2 (10.8)	Fasted
Cholesterol (mg dl ⁻¹)	135 (7)	183 (20)†	204 (16)†	Fasted
Insulin (ng ml ⁻¹)	1.77 (0.64) 0.73 (0.14)	9.21 (1.95)† 2.70 (0.36)†	2.46 (0.47)* 1.06 (0.30)*	Fed Fasted
Glucose (mg dl ⁻¹)	129.0 (5.4) 94.5 (3.3)	118.3 (4.7) 125.3 (11.6)†	114.8 (6.3) 85.6 (10.3)*	Fed Fasted
IGF-I (ng ml ⁻¹)	346 (40) 625 (33)	534 (12)† 999 (102)†	482 (21)†‡ 929 (81)†	Fed Fasted
IGFBP-1 (AU)	1.0 (0.3) 1.0 (0.2)	1.7 (0.3) 0.5 (0.3)	1.7 (1.0) 0.3 (0.1)†	Fed Fasted
IGFBP-2 (AU)	1.0 (0.2)	0.7 (0.04)	0.9 (0.1)	Fasted
Leptin (ng ml ⁻¹)	2.0 (1.1)	21.6 (7.2)	11.6 (6.5)	Fasted
Adiponectin (μg ml ⁻¹)	12.1 (1.6)	9.5 (0.5)	9.0 (0.8)	Fed
Amylase (U l ⁻¹)	2,060 (150)	2,960 (320)†	2,190 (230)*	Fasted
Ala aminotransferase (U l ⁻¹)	347 (119)	390 (61)	446 (88)	Fasted
Asp aminotransferase (U l ⁻¹)	448 (85)	425 (90)	512 (46)	Fasted
Creatine phosphokinase (U l ⁻¹)	4,260 (1820)	2,010 (810)	2,520 (680)	Fasted
Lactate dehydrogenase (U l ⁻¹)	1,530 (240)	1,610 (170)	2,020 (180)	Fasted
Alkaline phosphatase (U l ⁻¹)	43.8 (3.4)	44.6 (6.0)	34.2 (1.4)	Fasted
Bilirubin (mg dl ⁻¹)	0.16 (0.02)	0.10 (0.03)	0.16 (0.02)	Fasted
Albumin (g dl ⁻¹)	2.78 (0.16)	2.88 (0.19)	2.66 (0.14)	Fasted
Creatinine (mg dl ⁻¹)	0.54 (0.02)	0.48 (0.04)	0.46 (0.04)	Fasted
Cyclo-oxygenase (liver, AU mg ⁻¹)	1.00 (0.14)	0.80 (0.11)	0.83 (0.11)	Fed
Citrate synthase (liver, AU mg ⁻¹)	141 (14)	128 (21)	138 (11)	Fed
Body temperature (°C)	34.71 (0.14)	35.52 (0.17)†	35.57 (0.15)†	Fed

Values shown are mean (±s.e.m.). AU, arbitrary units; U l⁻¹, units per litre.

* $P < 0.05$ versus high calorie.

† $P < 0.05$ versus standard diet.

‡ $P < 0.05$ versus high calorie by one-tailed Student's *t*-test.

regeneration³⁹, is of note because its activity is known to be suppressed in the liver by high caloric diets and shows an age-related decline in activity that is attenuated by caloric restriction^{40,41}.

A few of the pathway changes were unanticipated. Although we had observed an increase in mitochondrial number in the HCR group, there was a decrease in the transcription of numerous mitochondrial genes, suggesting that the turnover of mitochondrial proteins was reduced. This result was unexpected, but is consistent with a previous report showing that SIRT1-dependent activation of PGC-1 α does not enhance transcription of mitochondrial genes²⁷. Upregulation of complement, which occurs in obese and aged mice, was also observed in the HCR group for reasons that are currently unclear.

It is notable that resveratrol opposed the effects of high caloric intake in 144 out of 153 significantly altered pathways (Fig. 4c). In fact, the PAGE signature of the HCR group was considerably more similar to that of the SD group than the HC controls. Principal component analysis yielded values of -1.82 (SD), -1.41 (HCR) and 3.22 (HC), with 88.4% of the variability assigned to the first principal component, making the HC group the clear outlier (Fig. 4d).

We next compared our PAGE results to a pre-existing caloric restriction data set for C57BL/6 mice known as AGEMAP, hypothesizing that the comparison of changes induced by these two paradigms

might reveal pathways common to the enhancement of health and longevity. Of the 36 different pathways identified by AGEMAP as being significantly altered by caloric restriction, there was sufficient overlap to compare 19 of them to our data (Fig. 4e). Pathways altered in the same direction by caloric restriction and resveratrol included the downregulation of IGF-1 and mTOR signalling, downregulation of glycolysis, and upregulation of Stat3 signalling. One interesting difference was that cell cycle checkpoint and apoptotic pathways were elevated in the caloric restriction group but downregulated by resveratrol. We do not favour the interpretation that the resveratrol-treated livers were undergoing less apoptosis because levels of AST and ALT, two indicators of hepatic apoptosis, were unchanged (see Table 1). Perhaps the downregulation of cell cycle checkpoints is linked to the recent discovery that inhibition of checkpoint function in *C. elegans* increases stress resistance and lifespan⁴². Although the statistical power of this analysis is limited by the overlap in data sets, the results suggest that more comprehensive comparisons of the effects of resveratrol and caloric restriction are warranted.

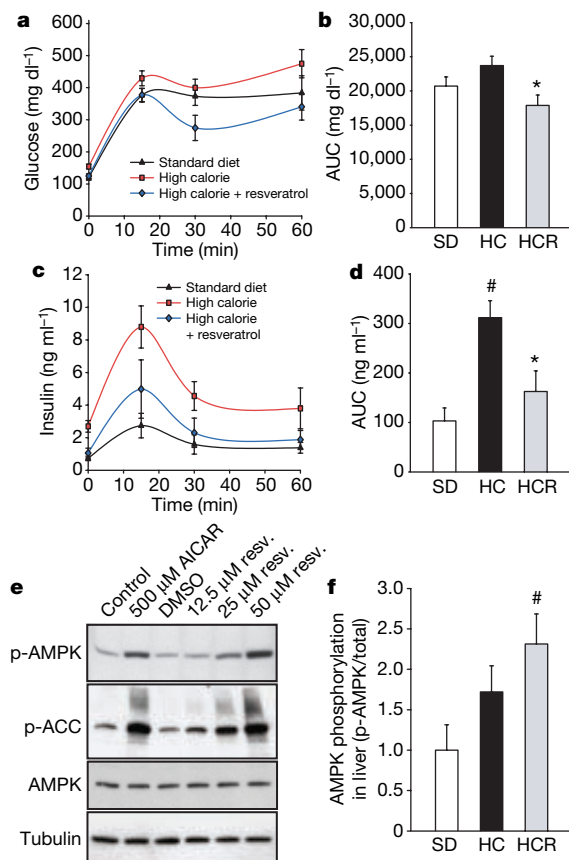


Figure 2 | Resveratrol improves insulin sensitivity and activates AMPK. **a–d**, Plasma levels of glucose (**a**, **b**) and insulin (**c**, **d**) were measured after a 2 g kg⁻¹ oral glucose dose. Areas under the curves (AUC) were significantly reduced by resveratrol treatment. **e**, Activation of AMPK by resveratrol in CHO cells. In the presence of resveratrol or 5-aminoimidazole-4-carboxamide-1- β -D-ribofuranoside (AICAR) as a positive control, phosphorylation of AMPK and its downstream target, acetyl-coA carboxylase (ACC), are increased. **f**, AMPK activity in liver. Phosphorylation of AMPK (**f**), acetyl-coA carboxylase (Supplementary Fig. 3e) and decreased expression of fatty acid synthase (Supplementary Fig. 3f) are indicative of enhanced AMPK activity. Asterisk, $P < 0.05$ versus HC; hash, $P < 0.05$ versus SD. $n = 5$ for all groups. Error bars indicate s.e.m.

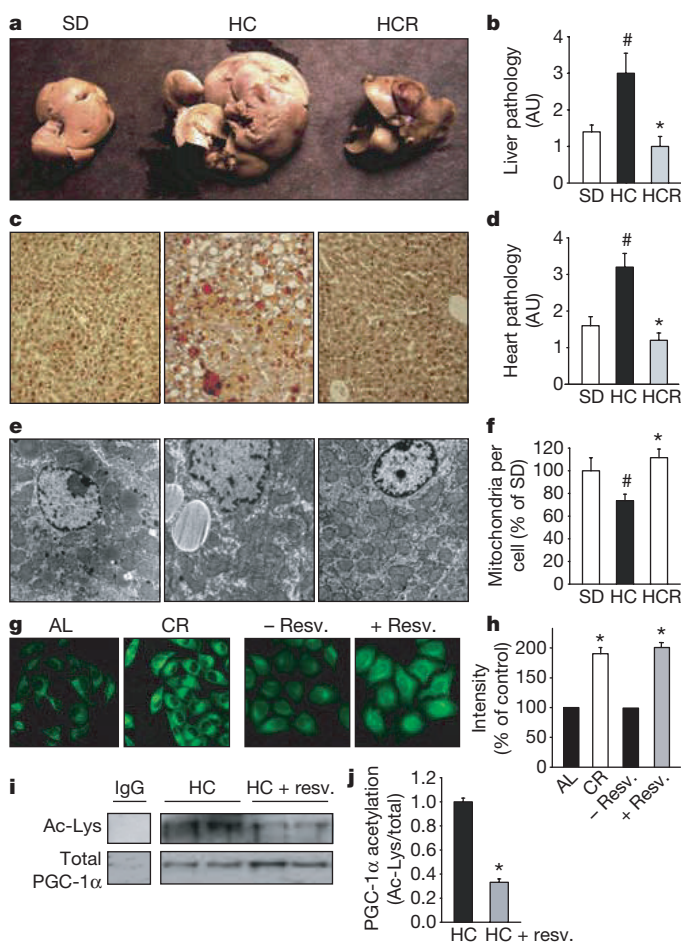


Figure 3 | Resveratrol improves liver histology, increases mitochondrial number and decreases acetylation of PGC-1 α . **a–c**, Resveratrol prevents the development of fatty liver, as assessed by organ size (**a**), overall pathology (**b**) and decreased fat accumulation as measured by oil red O staining (**c**). AU, arbitrary units. **d**, Pathology of heart sections. Additional histology of liver, heart and aorta is shown in Supplementary Fig. 4. **e**, **f**, Transmission electron microscopy of liver sections (**e**) and mitochondrial counts (**f**). **g**, **h**, Mitochondrial number in HeLa cells treated with serum from *ad libitum*-fed (AL) or calorically restricted (CR) rats, or resveratrol, and stained with Mitotracker green FM. **i**, **j**, Resveratrol reduces the acetylation of PGC-1 α , a known SIRT1 target and regulator of mitochondrial biogenesis, *in vivo*. PGC-1 α was immunoprecipitated from liver extracts then blotted for acetyl lysine (**i**) and quantified (**j**). Asterisk, $P < 0.05$ versus HC; hash, $P < 0.05$ versus SD. $n = 5$ for **b** and **d**; $n = 3$ for **f** and **j**. Error bars indicate s.e.m.

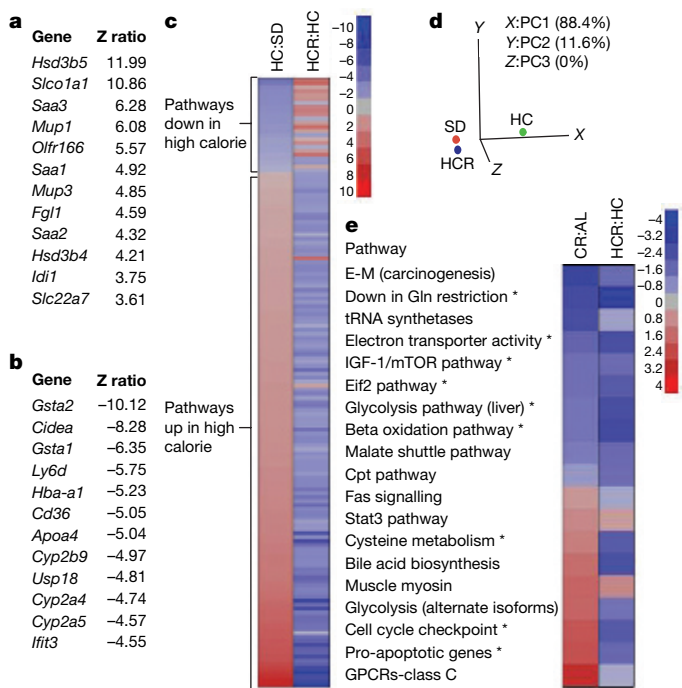


Figure 4 | Resveratrol shifts expression patterns of mice on a high-calorie diet towards those on a standard diet. a, b, The most highly significant upregulated (a) and downregulated (b) genes in livers of HC and HCR groups are shown. **c,** Parametric analysis of gene-set enrichment (PAGE) comparing every pathway significantly upregulated (red) or downregulated (blue) by either the HC diet or resveratrol (153 in total, with 144 showing opposing effects). **d,** Principal component analysis of PAGE data. The first principal component (PC1) is dominant, with 88.4% variability, and shows HCR to be more similar to SD than HC. **e,** Comparison of pathways significantly altered by resveratrol treatment and caloric restriction using data from the AGEMAP caloric restriction study. Pathways with significant differences between HC and HCR are indicated by an asterisk. Complete pathway listings are in Supplementary Fig. 7. Asterisk, $P < 0.05$ versus HC. $n = 5$ for SD and HC; $n = 4$ for HCR.

Discussion

The ability of resveratrol to prevent the deleterious effects of excess caloric intake and modulate known longevity pathways suggests that resveratrol and molecules with similar properties might be valuable tools in the search for key regulators of energy balance, health and longevity. As a case in point, the most highly upregulated gene in the HC group and second most highly downregulated gene in the HCR group was *Cidea*, which regulates energy balance in brown fat and provides resistance to obesity and diabetes when knocked out⁴³.

Taken together, the findings in this study show that resveratrol shifts the physiology of mice consuming excess calories towards that of mice on a standard diet, modulates known longevity pathways, and improves health, as indicated by a variety of measures including survival, motor function, insulin sensitivity, organ pathology, PGC-1 α activity, and mitochondrial number. Notably, all these changes occurred without a significant reduction in body weight. Whether these effects are due to resveratrol working primarily through SIRT1, which is the case for simpler metazoans, or through a combination of interactions, as predicted by the xenohormesis hypothesis^{6,44}, remains to be determined. In either case, this study shows that an orally available small molecule at doses achievable in humans can safely reduce many of the negative consequences of excess caloric intake, with an overall improvement in health and survival.

METHODS

Animals and diets. Animal housing and diets are described in Supplementary Information. Briefly, one-year-old male C57BL/6NIA mice were maintained on

AIN-93G standard diet (SD), AIN-93G modified to provide 60% of calories from fat (HC), or HC diet with the addition of 0.04% resveratrol (HCR).

Electron microscopy. Tissues were processed as previously reported, imaged at $\times 1,500$ on a Jeol 1210 transmission microscope, and photographed using a Gatan US 4000 MP Digital camera. Mitochondria per cell and hepatocyte size were quantified using grid counting by three blinded raters.

PAGE analysis. RNA extracted from the livers of five 18-month-old mice per group was hybridized to Agilent 44k whole-genome microarrays following protocols listed on the Gene Expression and Genomics Unit website at the National Institute on Aging (<http://www.grc.nia.nih.gov/branches/rrb/dna/index/protocols.htm>). One array from the HCR group was discarded owing to RNA degradation. Raw data were subjected to Z normalization and tested for significant changes as previously described²⁹. For parametric analysis of gene set enrichment (PAGE), a complete set of 522 pathways in the cell was obtained from http://www.broad.mit.edu/gsea/msigdb/msigdb_index.html. Details of the method are described in Supplementary Information and elsewhere³¹. Principal component analysis was performed on the replicate average for the three groups. These tools are part of DIANE 1.0, a program developed by V.V.P. and available at http://www.grc.nia.nih.gov/branches/rrb/dna/diane_software.pdf. Caloric restriction data were extracted from the AGEMAP project.

Cell-based mitochondrial assays. Mitochondrial mass was determined using the mitochondria-specific fluorescent dye Mitotracker green FM after fixation of cells, with minor modifications. Staining after fixation permits determination of the incorporation of the dye due to mitochondrial mass independently of the mitochondrial membrane potential ($\Delta\psi_m$).

Serum markers and hormones. Details of biochemical assays are described in Supplementary Information.

AMPK and PGC-1 α analysis. Extraction of tissues and cells for western blotting were performed using standard techniques. Details and antibodies are outlined in Supplementary Information.

Histology. Eighteen-month-old mice were fasted overnight and fixed using Streck fixative (Streck). Organs were sectioned and stained with haematoxylin and eosin or frozen and stained with oil red O lipid stain, then scored blindly for overall pathology.

Rotarod. Results shown are the average of three trials per mouse, measuring time to fall from an accelerating rotarod (4–40 r.p.m. over 5 min). Data from animals that survived to 24 months are shown.

Statistical analyses. Single factor analysis of variance followed by Fisher's post-hoc tests were used for all comparisons, except where noted. Details are provided in Supplementary Information.

Received 10 August; accepted 19 October 2006.

Published online 1 November 2006.

- Olshansky, S. J. Projecting the future of U.S. health and longevity. *Health Aff. (Millwood)* **24** (suppl. 2), W5R86–W5R89 (2005).
- Li, Z., Bowerman, S. & Heber, D. Health ramifications of the obesity epidemic. *Surg. Clin. North Am.* **85**, 681–701 (2005).
- Ingram, D. K. et al. Development of caloric restriction mimetics as a prolongevity strategy. *Ann. NY Acad. Sci.* **1019**, 412–423 (2004).
- Sinclair, D. A. Toward a unified theory of caloric restriction and longevity regulation. *Mech. Ageing Dev.* **126**, 987–1002 (2005).
- Guarente, L. & Picard, F. Caloric restriction—the SIR2 connection. *Cell* **120**, 473–482 (2005).
- Howitz, K. T. et al. Small molecule activators of sirtuins extend *Saccharomyces cerevisiae* lifespan. *Nature* **425**, 191–196 (2003).
- Wood, J. G. et al. Sirtuin activators mimic caloric restriction and delay ageing in metazoans. *Nature* **430**, 686–689 (2004).
- Viswanathan, M., Kim, S. K., Berdichevsky, A. & Guarente, L. A role for SIR-2.1 regulation of ER stress response genes in determining *C. elegans* life span. *Dev. Cell* **9**, 605–615 (2005).
- Jarolim, S. et al. A novel assay for replicative lifespan in *Saccharomyces cerevisiae*. *FEMS Yeast Res.* **5**, 169–177 (2004).
- Valenzano, D. R. et al. Resveratrol prolongs lifespan and retards the onset of age-related markers in a short-lived vertebrate. *Curr. Biol.* **16**, 296–300 (2006).
- Picard, F. et al. Sirt1 promotes fat mobilization in white adipocytes by repressing PPAR- γ . *Nature* **429**, 771–776 (2004).
- Chen, J. et al. SIRT1 protects against microglia-dependent amyloid- β toxicity through inhibiting NF- κ B signaling. *J. Biol. Chem.* **280**, 40364–40374 (2005).
- Kolthur-Seetharam, U., Dantzer, F., McBurney, M. W., de Murcia, G. & Sassone-Corsi, P. Control of AIF-mediated cell death by the functional interplay of SIRT1 and PARP-1 in response to DNA damage. *Cell Cycle* **5**, 873–877 (2006).
- Raval, A. P., Dave, K. R. & Perez-Pinzon, M. A. Resveratrol mimics ischemic preconditioning in the brain. *J. Cereb. Blood Flow Metab.* **26**, 1141–1147 (2006).
- Frescas, D., Valenti, L. & Accili, D. Nuclear trapping of the forkhead transcription factor FoxO1 via Sirt-dependent deacetylation promotes expression of glucogenic genes. *J. Biol. Chem.* **280**, 20589–20595 (2005).

16. Denu, J. M. The Sir2 family of protein deacetylases. *Curr. Opin. Chem. Biol.* **9**, 431–440 (2005).
17. Weindrich, R. & Walford, R. L. Dietary restriction in mice beginning at 1 year of age: effect on life-span and spontaneous cancer incidence. *Science* **215**, 1415–1418 (1982).
18. Siebler, J. & Galle, P. R. Treatment of nonalcoholic fatty liver disease. *World J. Gastroenterol.* **12**, 2161–2167 (2006).
19. Scrocchi, L. A. & Drucker, D. J. Effects of aging and a high fat diet on body weight and glucose tolerance in glucagon-like peptide-1 receptor^{-/-} mice. *Endocrinology* **139**, 3127–3132 (1998).
20. McCarty, M. F. Chronic activation of AMP-activated kinase as a strategy for slowing aging. *Med. Hypotheses* **63**, 334–339 (2004).
21. Apfeld, J., O'Connor, G., McDonagh, T., DiStefano, P. S. & Curtis, R. The AMP-activated protein kinase AAK-2 links energy levels and insulin-like signals to lifespan in *C. elegans*. *Genes Dev.* **18**, 3004–3009 (2004).
22. Zang, M. *et al.* Polyphenols stimulate AMP-activated protein kinase, lower lipids, and inhibit accelerated atherosclerosis in diabetic LDL receptor-deficient mice. *Diabetes* **55**, 2180–2191 (2006).
23. Nisoli, E. *et al.* Calorie restriction promotes mitochondrial biogenesis by inducing the expression of eNOS. *Science* **310**, 314–317 (2005).
24. Lopez-Lluch, G. *et al.* Calorie restriction induces mitochondrial biogenesis and bioenergetic efficiency. *Proc. Natl Acad. Sci. USA* **103**, 1768–1773 (2006).
25. Wu, Z. *et al.* Mechanisms controlling mitochondrial biogenesis and respiration through the thermogenic coactivator PGC-1. *Cell* **98**, 115–124 (1999).
26. Yoon, J. C. *et al.* Control of hepatic gluconeogenesis through the transcriptional coactivator PGC-1. *Nature* **413**, 131–138 (2001).
27. Rodgers, J. T. *et al.* Nutrient control of glucose homeostasis through a complex of PGC-1 α and SIRT1. *Nature* **434**, 113–118 (2005).
28. Lerin, C. *et al.* GCN5 acetyltransferase complex controls glucose metabolism through transcriptional repression of PGC-1 α . *Cell Metab.* **3**, 429–438 (2006).
29. Cheadle, C., Vawter, M. P., Freed, W. J. & Becker, K. G. Analysis of microarray data using Z score transformation. *J. Mol. Diagn.* **5**, 73–81 (2003).
30. Baur, J. A. & Sinclair, D. A. Therapeutic potential of resveratrol: the *in vivo* evidence. *Nature Rev. Drug Discov.* **5**, 493–506 (2006).
31. Kim, S. Y. & Volsky, D. J. PAGE: parametric analysis of gene set enrichment. *BMC Bioinformatics* **6**, 144 (2005).
32. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* **102**, 15545–15550 (2005).
33. Feng, J., Bussiere, F. & Hekimi, S. Mitochondrial electron transport is a key determinant of life span in *Caenorhabditis elegans*. *Dev. Cell* **1**, 633–644 (2001).
34. Kenyon, C. A conserved regulatory mechanism for aging. *Cell* **105**, 165–168 (2001).
35. Kapahi, P. *et al.* Regulation of lifespan in *Drosophila* by modulation of genes in the TOR signaling pathway. *Curr. Biol.* **14**, 885–890 (2004).
36. Kaerberlein, M. *et al.* Regulation of yeast replicative life span by TOR and Sch9 in response to nutrients. *Science* **310**, 1193–1196 (2005).
37. Ingram, D. K. *et al.* Calorie restriction mimetics: an emerging research field. *Aging Cell* **5**, 97–108 (2006).
38. Hipkiss, A. R. Does chronic glycolysis accelerate aging? Could this explain how dietary restriction works? *Ann. NY Acad. Sci.* **1067**, 361–368 (2006).
39. Taub, R. Liver regeneration: from myth to mechanism. *Nature Rev. Mol. Cell Biol.* **5**, 836–847 (2004).
40. Xu, X. & Sonntag, W. E. Growth hormone-induced nuclear translocation of Stat-3 decreases with age: modulation by caloric restriction. *Am. J. Physiol.* **271**, E903–E909 (1996).
41. Wilsey, J. & Scarpese, P. J. Caloric restriction reverses the deficits in leptin receptor protein and leptin signaling capacity associated with diet-induced obesity: role of leptin in the regulation of hypothalamic long-form leptin receptor expression. *J. Endocrinol.* **181**, 297–306 (2004).
42. Olsen, A., Vantipalli, M. C. & Lithgow, G. J. Checkpoint proteins control survival of the postmitotic cells in *Caenorhabditis elegans*. *Science* **312**, 1381–1385 (2006).
43. Zhou, Z. *et al.* Cidea-deficient mice have lean phenotype and are resistant to obesity. *Nature Genet.* **35**, 49–56 (2003).
44. Howitz, K. T. & Sinclair, D. A. in *Handbook of the Biology of Aging* (eds Masoro, E. J. & Austad, S. N.) 63–104 (Elsevier, Boston, 2006).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank H. Rasnow for his donation towards purchasing the mice; V. Massari for help with electron microscopy; W. Matson for HPLC assays of resveratrol; M. Chachich and M. Tatar for help with statistical analysis; W. Wood for microarray assistance; D. Phillips for animal care; J. Egan's group for insulin measurements; the Sinclair laboratory, N. Wolf, P. Elliott, C. Westphal, R. Weindrich, P. Lambert, J. Milne and M. Milburn for advice on experimental design; and S. Luikenhuis and M. Dipp for critical reading of the manuscript. D.A.S. was supported by The Ellison Medical Research Foundation, J.A.B. by The American Heart Foundation, H.J. and D.L.C. by an Australian NHMRC project grant and an NHMRC post-graduate scholarship, and C.L. and P.P. by the American Diabetes Association and NIH. This research was supported (in part) by the Intramural Research Program of the NIH, NIA, a Spanish MCyT grant to P.N., NIH grants to D.A.S., and by the support of P. F. Glenn and The Paul F. Glenn Laboratories for the Biological Mechanisms of Aging.

Author Contributions All experiments were designed and carried out by J.A.B., K.J.P., R.deC. and D.A.S. with the exceptions noted below. D.K.I. assisted with the design and interpretation of experiments. N.L.P., A.K., J.S.A., K.L. and P.J.P. assisted with the handling of animals and behavioural studies. Electron microscopy was performed by H.A.J. and D.L.C. Computational methods for analysis of microarray data were developed and applied by V.V.P. and K.G.B. Analysis of PGC-1 α acetylation was performed by C.L. and P.P. Pathological assessments were made by S.P. (livers), M.W. and E.G.L. (hearts and aortas). *In vitro* AMPK reactions were performed by D.G. and R.J.S. Citrate synthase activity was measured by O.B. Cell-based mitochondrial assays were performed by G.L.L. and P.N. MRI images were obtained by S.R., K.W.F. and R.G.S.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details accompany the paper on www.nature.com/nature. Correspondence and requests for materials should be addressed to D.S. (david_sinclair@hms.harvard.edu) or R.deC. (deCaboRa@grc.nia.nih.gov).

Hydrodynamic turbulence cannot transport angular momentum effectively in astrophysical disks

Hantao Ji¹, Michael Burin¹†, Ethan Scharfman¹ & Jeremy Goodman¹

The most efficient energy sources known in the Universe are accretion disks. Those around black holes convert 5–40 per cent of rest-mass energy to radiation. Like water circling a drain, inflowing mass must lose angular momentum, presumably by vigorous turbulence in disks, which are essentially inviscid¹. The origin of the turbulence is unclear. Hot disks of electrically conducting plasma can become turbulent by way of the linear magnetorotational instability². Cool disks, such as the planet-forming disks of proto-stars, may be too poorly ionized for the magnetorotational instability to occur, and therefore essentially unmagnetized and linearly stable. Nonlinear hydrodynamic instability often occurs in linearly stable flows (for example, pipe flows) at sufficiently large Reynolds numbers. Although planet-forming disks have extreme Reynolds numbers, keplerian rotation enhances their linear hydrodynamic stability, so the question of whether they can be turbulent and thereby transport angular momentum effectively is controversial^{3–15}. Here we report a laboratory experiment, demonstrating that non-magnetic quasi-keplerian flows at Reynolds numbers up to millions are essentially steady. Scaled to accretion disks, rates of angular momentum transport lie far below astrophysical requirements. By ruling out purely hydrodynamic turbulence, our results indirectly support the magnetorotational instability as the likely cause of turbulence, even in cool disks.

Our experiments involved a novel Taylor–Couette apparatus¹⁶. The rotating liquid (water or a water/glycerol mixture) is confined between two concentric cylinders of radii r_1 , r_2 ($r_2 > r_1$) and height h . The angular velocity of the fluid is controlled by that of the cylinders, Ω_1 and Ω_2 . An infinitely long, steady, Taylor–Couette flow rotates as:

$$\Omega(r) = a + \frac{b}{r^2} \quad (1)$$

where $a = (\Omega_2 r_2^2 - \Omega_1 r_1^2) / (r_2^2 - r_1^2)$ and $b = r_1^2 r_2^2 (\Omega_1 - \Omega_2) / (r_2^2 - r_1^2)$. Reynolds number here can be defined as $\bar{r}(r_2 - r_1)(\Omega_1 - \Omega_2) / \nu$, where ν is viscosity and $\bar{r} \equiv (r_1 + r_2)/2$. The rotation profile (equation (1)) ensures a radially constant viscous torque $-2\pi\rho\nu h r^3 \partial\Omega/\partial r$ for constant mass density ρ and constant ν . Astrophysical disks are mostly keplerian, meaning $\Omega \propto r^{-3/2}$, so that $|\Omega|$ decreases radially outward ($\partial|\Omega|/\partial r < 0$) while the specific angular momentum, $|r^2\Omega|$, increases radially ($\partial|r^2\Omega|/\partial r > 0$). We apply the term ‘quasi-keplerian’ to any flow satisfying these conditions, which are crucial for both hydrodynamic and magnetohydrodynamic linear stability².

Real flows have finite length. Disks have nearly stress-free vertical boundaries, but viscous stress at the vertical endcaps of laboratory flows drives secondary circulation. This may cause the rotation profile to deviate significantly from equation (1), and may even provoke turbulence^{7,15,17–20}. Our apparatus incorporates two rings at each end that are driven independently of the cylinders (Fig. 1). Thus we have four controllable angular velocities. When we choose these

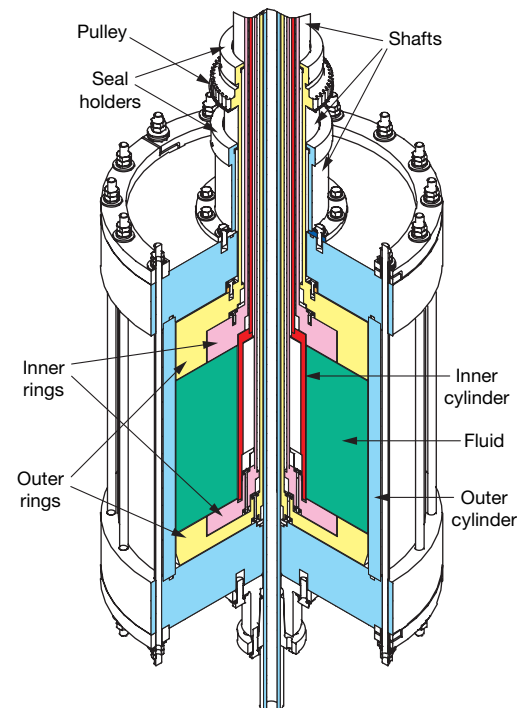


Figure 1 | Experimental set-up. A rotating fluid (water or a water/glycerol mixture) of height $h = 27.86$ cm is confined between two concentric cylinders of radii $r_1 = 7.06$ cm and $r_2 = 20.30$ cm, which rotate at rates of Ω_1 and Ω_2 , respectively. Two novel features distinguish this apparatus from conventional Taylor–Couette experiments. First, secondary circulation is controlled by dividing each endcap into two independently driven rings. Opposing rings at top and bottom are driven at the same selectable angular velocity Ω_3 (inner rings) and Ω_4 (outer rings). Traditionally, a large aspect ratio $\Gamma \equiv h/(r_2 - r_1)$ is used to reduce the secondary circulation. However, at $\Gamma \approx 25$ by Richard⁷ and even at $\Gamma \approx 100$ by Taylor¹⁸, end effects were reported to be significant when the endcaps co-rotated with one of the cylinders. Even when the endcaps were divided into two rings, but with each affixed to one cylinder^{7,17}, residual secondary circulation may have facilitated the observed turbulent transition^{15,19}. When Ω_3 and Ω_4 are appropriately chosen, secondary circulation is minimized and ideal Couette profiles are well approximated²¹. A second novel feature is access to rotation profiles on both sides of marginal linear stability at Reynolds numbers as large as 10^6 (see also Fig. 2). When the specific angular momentum, $|r^2\Omega|$, decreases with increasing r , the Rayleigh stability criterion³⁰ is violated, and thus the flow is linearly unstable when the Reynolds number exceeds a critical value¹⁶. When $\partial|r^2\Omega|/\partial r > 0$ but $\partial|\Omega|/\partial r < 0$ (as in disks, where $\Omega \propto r^{-3/2}$), the flow is quasi-keplerian and known to be linearly stable to axisymmetric disturbances. All major components of the apparatus were precisely machined and balanced, and except for the inner cylinder and rotating shafts, are made of clear acrylic to facilitate visual and laser diagnostics.

¹Center for Magnetic Self-organization in Laboratory and Astrophysical Plasmas, Plasma Physics Laboratory and Department of Astrophysical Sciences, Princeton University, Princeton, New Jersey 08543, USA. †Present address: Department of Physics and Astronomy, Pomona College, Claremont, California 91711, USA.

appropriately, secondary circulation is minimized, and ideal Couette profiles are closely approximated throughout the flow, except within ~ 1 cm of the endcaps²¹.

Most past work has taken the outer cylinder at rest ($\Omega_2 = 0$), so that both $|\Omega|$ and $|r^2\Omega|$ decrease radially outward. Such flows are axisymmetrically linearly unstable¹⁶ at modest Reynolds number, Re . Very few experiments have studied the linearly stable regime where $\partial|r^2\Omega|/\partial r > 0$, as occurs in disks. Among these few are two classic experiments of the 1930s: in one of these, the inner cylinder was at rest ($\Omega_1 = 0$)¹⁸, while in the other¹⁷, $0 \leq \Omega_1/\Omega_2 \leq 1$. Enhanced torques between the cylinders and other evidence of turbulence were reported at sufficiently large Re . These results have been cited in support of the hypothesis of nonlinear hydrodynamic turbulent transport in disks^{3,6}, notwithstanding that the direction of turbulent angular momentum transport, which always follows $-\partial\Omega/\partial r$ on energetic grounds, differs in sign between these experiments and astrophysical disks. The so-called β prescription⁶ derived from the above experiments has been used to model or interpret astronomical observations²². To our knowledge, the only published experiments with quasi-keplerian flow at relevant Reynolds numbers are those of Richard^{7,14} and Beckley²³. In Richard's work, transition to turbulence via nonlinear instabilities was studied qualitatively by a flow-visualization method. No direct measurements of angular momentum transport were performed. Beckley did measure torques roughly

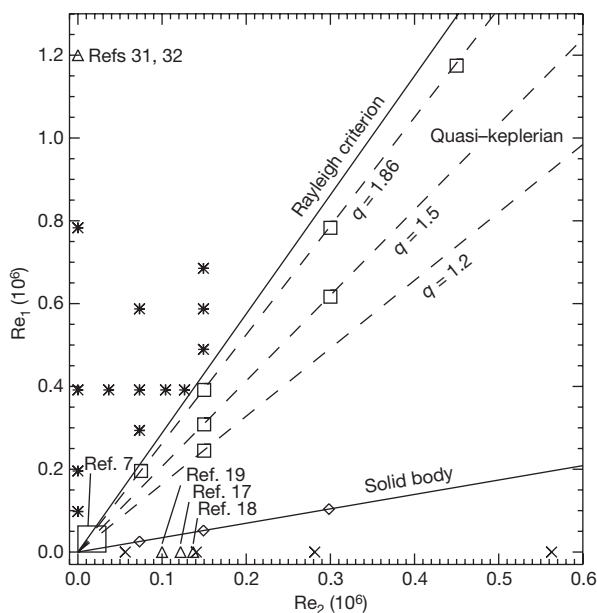


Figure 2 | Experimentally studied Taylor-Couette flows. Axes are Reynolds numbers based on the inner and outer cylinders: $Re_{1,2} \equiv \Omega_{1,2} r_{1,2} (r_2 - r_1) / \nu$. Asterisks mark Rayleigh-unstable flows; squares, quasi-keplerian ones, that is $\partial|\Omega|/\partial r < 0$ but $\partial|r^2\Omega|/\partial r > 0$; diamonds, solid-body flows ($\partial\Omega/\partial r = 0$); crosses for the inner cylinder at rest; triangles for flows explored in previous experiments^{17–19,31,32}; and the rectangular box for the parameter regime explored by Richard⁷. Dashed lines denote constant values of $q \equiv -\partial \ln \Omega / \partial \ln r$ at $r = 17$ cm, where most of our measurements of Reynolds stress were performed (Fig. 3). Rayleigh-unstable flows exhibit 5–10% fluctuations that are insensitive to the end-ring speeds. Quasi-keplerian and other Rayleigh-stable flows are more sensitive. For example, when the end-rings are fixed to the cylinders, fluctuations up to 4–8% occur. When the end-ring speeds are adjusted so that the ideal-Couette profile is restored, fluctuations are 1–2% and indistinguishable from those of our solid-body flows, except within a few cm of the boundary. Reducing Re by a factor of $\sim 1/18$, using an admixture of glycerol, actually increases the fluctuation level for the same (nearly ideal-Couette) profile. We interpret this to mean that the residual unsteady secondary circulation penetrates deeper into the bulk flow at lower Re . We infer from this that the experiments by Richard⁷ may have been affected by the endcaps, although his ratio of height to gap width exceeded ours. His endcaps were split but fixed to the cylinders.

consistent with the β prescription, but attributed them to secondary circulation, which was strong because the endcaps of his apparatus co-rotated with the outer cylinder and $h/(r_2 - r_1)$ was only ~ 2 .

Experimental flows studied by ourselves and others are summarized in Fig. 2. Our Reynolds numbers are up to 20 times larger than those previously achieved by Richard⁷. A laser Doppler velocimetry (LDV) system was used to sample the azimuthal velocity v_θ at various radial and axial locations. Both mean values, $\bar{v}_\theta \equiv r\Omega$, and fluctuations, $v'_\theta \equiv v_\theta - \bar{v}_\theta$, were obtained. At our Reynolds numbers, linearly-unstable flows are always turbulent, with fluctuation levels $(\overline{v'^2_\theta})^{1/2} / \bar{v}_\theta = 5 - 10\%$. They are largely insensitive to the endcap speeds. In contrast, quasi-keplerian and other linearly-stable flows are sensitive to the endcap boundary conditions. When the endcap speeds are adjusted to best approximate ideal Couette flow, the fluctuations are 1–2% and indistinguishable from those of our solid-body flows, which are expected to be steady or laminar owing to the lack of shear to drive turbulence.

One hallmark of nonlinear instability is hysteresis: the transition from laminar flow to turbulence occurs at higher Reynolds numbers than the reverse⁷. Quiescent flows were gradually brought into linearly-unstable regimes by raising Ω_1 or lowering Ω_2 , then returned to linearly-stable regimes. Significant fluctuations were found only in the linearly-unstable regime; no hysteresis was detected.

In the absence of magnetic fields, turbulent angular momentum transport requires correlated velocity fluctuations. The radial angular momentum flux is $\rho r \bar{v}'_\theta v'_r$, where v'_r is the fluctuation in radial velocity, and the turbulent viscosity ν_{turb} is defined by equating this to $-\rho \nu_{\text{turb}} r^2 \partial\Omega/\partial r$ where $\nu_{\text{turb}} = \beta |r^2 \partial\Omega/\partial r|$, so that $\beta = \overline{v'_\theta v'_r} / (r^2 \partial\Omega/\partial r)^2$ is dimensionless. Conveniently, in a turbulent but statistically steady state with a profile given by equation (1), both β and ν_{turb} are radially constant. A value of $\beta = (1-2) \times 10^{-5}$ has been deduced⁶ from experiments^{17,18} with $0 \leq \Omega_1/\Omega_2 < 1$.

We have measured the Reynolds stress directly using a synchronized dual-LDV system. Both v'_θ and v'_r appear to follow gaussian statistics (E.S. *et al.*, manuscript in preparation), and random errors

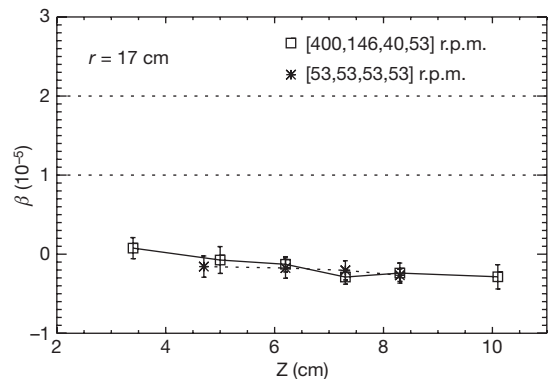


Figure 3 | Experimentally measured Reynolds stress versus height in a quasi-keplerian profile. Here $\beta \equiv \overline{v'_\theta v'_r} / (\overline{v'^2_\theta} q^2)$ sign $(q\Omega)$, where $q \equiv -\partial \ln \Omega / \partial \ln r$: $q = +3/2$ in keplerian disks, $1.2 \leq q \leq 1.9$ in our quasi-keplerian flows. Square symbols connected by a solid line were taken at $q = 1.86$ (see Fig. 2). Starred points connected by a dotted line are a solid-body case ($q = 0$), which should be non-turbulent and therefore serves as a control for systematic errors. The measured values fall far below the range of β proposed by Richard and Zahn⁶, shown with horizontal dotted lines. The measurements were performed using a synchronized, dual laser-Doppler-velocimetry (LDV) system, which allows simultaneous detection of both components of velocity v_θ and v_r . The laser beams enter the fluid vertically through the acrylic endcaps from below (see Fig. 1), and Z is the height above the lower endcap. Rotation speeds $[\Omega_1, \Omega_3, \Omega_4, \Omega_2]$ are shown, where Ω_3 and Ω_4 are the angular velocities of the inner and outer end-rings. Apparently gaussian deviations amounting to $\sim (1-2\%)$ of $\bar{v}_\theta$ were observed from sample to sample, representing measurement errors and perhaps true fluctuations. Each data point is the mean of β computed from 3,000–10,000 samples. Error bars represent the standard deviations.

were reduced by averaging 10^3 – 10^4 samples. Systematic errors were gauged by comparison with solid body flows ($\Omega_1 = \Omega_2 = \Omega_3 = \Omega_4$). Figure 3 shows that β differs indistinguishably between quasi-keplerian and solid-body rotation, falling far below the proposed range⁶. A large outward Reynolds stress is detected in linearly-unstable flows, $\beta > 10^{-3}$, as expected (M.B. *et al.*, manuscript in preparation). Even for linearly-stable flows, when the speeds (Ω_3, Ω_4) of the endcaps were not adjusted properly to produce the ideal Couette profile of equation (1), β was found to be almost 10^{-4} (Fig. 4). This again indicates that the axial boundaries can profoundly influence linearly-stable flows, as previously suggested^{15,19}.

A final point should be made about the critical Reynolds number for transition, Re_{crit} , versus gap width. Based on the experiments reported in refs 17 and 18, the scaling $Re_{crit} \approx 6 \times 10^5 [(r_2 - r_1)/\bar{r}]^2$ has been proposed^{3,4,6,8}. It is unclear whether this scaling applies to quasi-keplerian flow as it was derived from data for $\Omega_2/\Omega_1 > 1$. In any case, up to $Re = 2 \times 10^6$, which is about three times the proposed

Re_{crit} because $(r_2 - r_1)/\bar{r} \approx 1$ in our device, we have seen no signs of rising fluctuation levels or Reynolds stress (Fig. 4).

Therefore, we have shown that purely hydrodynamic quasi-keplerian flows, under proper boundary conditions and at large enough Reynolds numbers, cannot transport angular momentum at astrophysically relevant rates.

Of course, it could be argued that our maximum Re , which only barely exceeds some theoretical estimates¹³ of Re_{crit} , is still not large enough for transition. Or it could be that transition has occurred, but that the transport is too small for us to detect. To extrapolate from $Re \leq 2 \times 10^6$ to a typical astrophysical value $\gtrsim 10^{12}$, we rely on the empirical observation that for $Re > Re_{crit}$, the Reynolds number based on the turbulent rather than the molecular viscosity is approximately independent of Re itself. It follows that $v_{turb} \approx LU/Re_{crit}$ for $Re > Re_{crit}$, where L and U are the characteristic size and velocity of the flow. It is common knowledge among civil engineers that this is true of flow in pipes, for example²⁴. If it is true of rotating shear flow, as theoretical arguments suggest it should be¹⁵, then β at $Re \gtrsim 10^{12}$ should be comparable to what we find at $Re \lesssim 2 \times 10^6$, namely, $\beta < 6.2 \times 10^{-6}$ (at 2 s.d., or 98% confidence; see Fig. 4 legend)—whether or not we have crossed the threshold of transition.

Lastly, it is useful to relate the above upper bound for β to the more commonly used Shakura–Sunyaev α parameter¹: $v_{turb} = \alpha \Omega h^2$. We replace this by $v_{turb} = \alpha \Omega (r_2 - r_1)^2$ since $r_2 - r_1$ is smaller than h in our experiment, and we presume that the dominant turbulent eddies scale with the smallest dimension of the flow ($h \ll r$ in most disks). Then $\alpha = \beta q \bar{r}^2 / (r_2 - r_1)^2 \approx \beta q$ (see Fig. 3 legend for a definition of q), and thus our results imply a similar upper bound for α as for β in purely hydrodynamic disks, whereas protostellar-disk lifetimes and accretion rates indicate²⁵ $\alpha \gtrsim 10^{-3}$. Although it has been suggested that complications such as vertical or radial stratification may yet lead to essentially linear non-axisymmetric hydrodynamic instabilities^{26,27}, our belief is that such non-axisymmetric linear instabilities depend upon radial boundaries and hence are not generally important in thin disks^{28,29}. If this is correct, then by default, the magnetorotational instability appears to be the only plausible source of accretion disk turbulence.

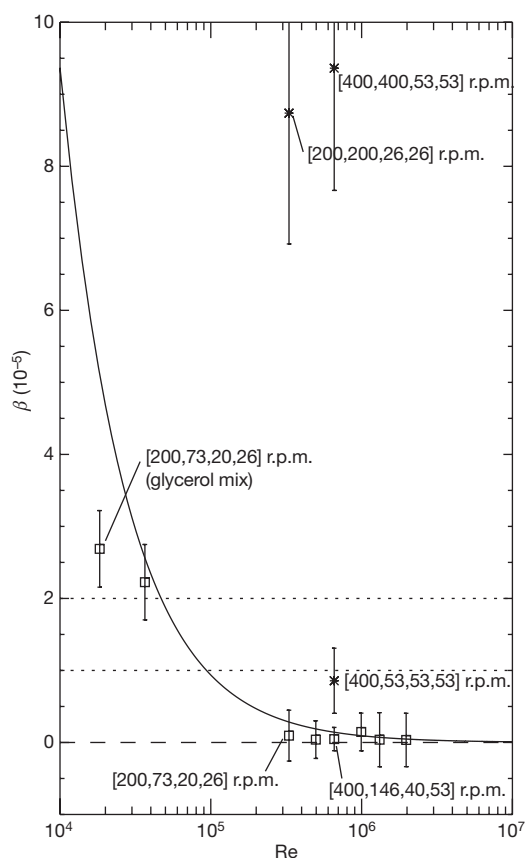


Figure 4 | Dimensionless Reynolds stress at Reynolds numbers up to 2×10^6 . At the largest Re , the data show no sign that $\beta \equiv v'v' / (\bar{v}_\theta^2 q^2)$ or the relative fluctuations $v'^2 / (\bar{v}_\theta)^2$ themselves increase with Re in quasi-keplerian flows. Here $q \equiv -\partial \ln \Omega / \partial \ln r$ is calculated from the mean $v_\theta(r)$ profiles. Systematic errors in β have been removed by reference to identical measurements in solid-body flows. Different sizes of error bars (showing s.d.) for the six quasi-keplerian flows with optimum boundary conditions (squares at $Re > 10^5$) are largely due to different numbers of measurement samples, N . Averaging over these 6 points, weighted by $\sqrt{N}$, results in $\beta = 0.72 \times 10^{-6}$ with s.d. of 2.7×10^{-6} or $\beta < 6.2 \times 10^{-6}$ at 2 s.d. (98% confidence). When the end-ring speeds are not optimized to produce Couette profiles (equation (1)), however, β exhibits large values (asterisks). When optimal speeds are used but glycerol is added to the water to reduce the Reynolds number, larger β values are also seen (squares at $Re < 10^5$), possibly due to stronger residual secondary circulation. For reference, the solid line is the molecular viscous value: $\beta_{visc} \equiv \nu / [\bar{r}^3 (\Omega_2 - \Omega_1) / (r_2 - r_1)] = Re^{-1} (r_2 - r_1)^2 / \bar{r}^2$. The proposed β lies between the horizontal dotted lines⁶.

Received 8 August; accepted 6 October 2006.

- Shakura, N. I. & Sunyaev, R. A. Black holes in binary systems. Observational appearance. *Astron. Astrophys.* **24**, 337–355 (1973).
- Balbus, S. A. & Hawley, J. F. Instability, turbulence, and enhanced transport in accretion disks. *Rev. Mod. Phys.* **70**, 1–53 (1998).
- Zeldovich, Y. B. On the friction of fluids between rotating cylinders. *Proc. R. Soc. Lond. A* **374**, 299–312 (1981).
- Dubrulle, B. Differential rotation as a source of angular momentum transfer in the solar nebula. *Icarus* **106**, 59–76 (1993).
- Balbus, S. A., Hawley, J. F. & Stone, J. M. Nonlinear stability, hydrodynamical turbulence, and transport in disks. *Astrophys. J.* **467**, 76–86 (1996).
- Richard, D. & Zahn, J.-P. Turbulence in differentially rotating flows: What can be learned from the Couette–Taylor experiment. *Astron. Astrophys.* **347**, 734–738 (1999).
- Richard, D. *Instabilités Hydrodynamiques dans les Écoulements en Rotation Différentielle*. Ph.D. thesis, Univ. Paris 7 (2001).
- Longaretti, P. On the phenomenology of hydrodynamic shear turbulence. *Astrophys. J.* **576**, 587–598 (2002).
- Chagelishvili, G. D., Zahn, J.-P., Tevzadze, A. G. & Lominadze, J. G. On hydrodynamic shear turbulence in Keplerian disks: Via transient growth to bypass transition. *Astron. Astrophys.* **402**, 401–407 (2003).
- Yecko, P. A. Accretion disk instability revisited. Transient dynamics of rotating shear flow. *Astron. Astrophys.* **425**, 385–393 (2004).
- Umurhan, O. M. & Regev, O. Hydrodynamic stability of rotationally supported flows: Linear and nonlinear 2D shearing box results. *Astron. Astrophys.* **427**, 855–872 (2004).
- Garaud, P. & Ogilvie, G. I. A model for the nonlinear dynamics of turbulent shear flows. *J. Fluid Mech.* **530**, 145–176 (2005).
- Mukhopadhyay, B., Afshordi, N. & Narayan, R. Bypass to turbulence in hydrodynamic accretion disks: An eigenvalue approach. *Astrophys. J.* **629**, 383–396 (2005).
- Dubrulle, B. *et al.* Stability and turbulent transport in Taylor–Couette flow from analysis of experimental data. *Phys. Fluids* **17**, 095103, doi:10.1063/1.2008999 (2005).

15. Lesur, G. & Longaretti, P.-Y. On the relevance of subcritical hydrodynamic turbulence to accretion disk transport. *Astron. Astrophys.* **444**, 25–44 (2005).
16. Taylor, G. I. Stability of a viscous liquid contained between two rotating cylinders. *Phil. Trans. R. Soc. Lond. A* **223**, 289–343 (1923).
17. Wendt, F. Turbulente Strömungen zwischen zwei rotierenden konaxialen Zylindern. *Ing. Arch.* **4**, 577–595 (1933).
18. Taylor, G. I. Fluid friction between rotating cylinders. i. torque measurements. *Proc. R. Soc. Lond. A* **157**, 546–578 (1936).
19. Schultz-Grunow, F. Zur Stabilität der Couette-Strömung. *Z. Angew. Math. Mech.* **39**, 101–117 (1959).
20. Kageyama, A., Ji, H., Goodman, J., Chen, F. & Shoshan, E. Numerical and experimental investigation of circulation in short cylinders. *J. Phys. Soc. Jpn* **73**, 2424–2437 (2004).
21. Burin, M. J. et al. Reduction of Ekman circulation within a short circular couette flow. *Exp. Fluids* **40**, 962–966, doi:10.1007/s00348-006-0132-y (2006).
22. Hueso, R. & Guillot, T. Evolution of protoplanetary disks: constraints from DM Tauri and GM Aurigae. *Astron. Astrophys.* **442**, 703–725 (2005).
23. Beckley, H. *Measurements of Annular Couette Flow Stability at the Fluid Reynolds Number $Re = 4.4 \times 10^6$: The Fluid Dynamic Precursor to a Liquid Sodium $\alpha\omega$ Dynamo*. PhD thesis, New Mexico Inst. Mining Technol. (2002).
24. Colebrook, C. F. Turbulent flow in pipes with particular reference to the transitional region between smooth and rough pipes. *J. Inst. Civil Eng.* **11**, 133–156 (1938).
25. Hartmann, L., Calvet, N., Gullbring, E. & D'Alessio, P. Accretion and the evolution of T Tauri disks. *Astrophys. J.* **495**, 385–400 (1998).
26. Klahr, H. H. & Bodenheimer, P. Turbulence in accretion disks: Vorticity generation and angular momentum transport via the global baroclinic instability. *Astrophys. J.* **582**, 869–892 (2003).
27. Dubrulle, B. et al. An hydrodynamic shear instability in stratified disks. *Astron. Astrophys.* **429**, 1–13 (2005).
28. Johnson, B. M. & Gammie, C. F. Nonlinear stability of thin, radially-stratified disks. *Astrophys. J.* **636**, 63–74 (2006).
29. Goodman, J. & Balbus, S. A. Stratified disks are locally stable. Preprint at (<http://arxiv.org/astro-ph/0110229>) (2001).
30. Rayleigh, Lord On the dynamics of rotating fluid. *Proc. R. Soc. Lond. A* **93**, 148–154 (1916).
31. Lathrop, D. P., Fineberg, J. & Swinney, H. L. Turbulent flow between concentric rotating cylinders at large Reynolds number. *Phys. Rev. Lett.* **68**, 1515–1518 (1992).
32. Lewis, G. S. & Swinney, H. L. Velocity structure functions, scaling, and transitions in high-Reynolds-number Couette-Taylor flow. *Phys. Rev. E* **59**, 5457–5467 (1999).

Acknowledgements We thank S. Balbus for discussions, R. Cutler for technical assistance with the apparatus, P. Heitzenroeder, C. Jun, L. Morris and S. Raftopolous for engineering assistance, as well as Dantec Dynamics for the contracted use of an LDV measurement system. This research was supported by the US Department of Energy, Office of Science – Fusion Energy Sciences Program; the US National Aeronautics and Space Administration, Astronomy and Physics Research and Analysis and Astrophysics Theory Programs; and the US National Science Foundation, Physics and Astronomical Sciences Divisions.

Author Contributions H.J., M.B. and E.S. planned and executed the experiments, and analysed data; E.S. and M.B. prepared apparatus and diagnostics; H.J. drafted the paper; and J.G. suggested this subject and assisted in the interpretation of the results and in revising the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to H.J. (hji@pppl.gov).

Half-metallic graphene nanoribbons

Young-Woo Son^{1,2}, Marvin L. Cohen^{1,2} & Steven G. Louie^{1,2}

Electrical current can be completely spin polarized in a class of materials known as half-metals, as a result of the coexistence of metallic nature for electrons with one spin orientation and insulating nature for electrons with the other. Such asymmetric electronic states for the different spins have been predicted for some ferromagnetic metals—for example, the Heusler compounds¹—and were first observed in a manganese perovskite². In view of the potential for use of this property in realizing spin-based electronics, substantial efforts have been made to search for half-metallic materials^{3,4}. However, organic materials have hardly been investigated in this context even though carbon-based nanostructures hold significant promise for future electronic devices⁵. Here we predict half-metallicity in nanometre-scale graphene ribbons by using first-principles calculations. We show that this phenomenon is realizable if in-plane homogeneous electric fields are applied across the zigzag-shaped edges of the graphene nanoribbons, and that their magnetic properties can be controlled by the external electric fields. The results are not only of scientific interest in the interplay between electric fields and electronic spin degree of freedom in solids^{6,7} but may also open a new path to explore spintronics³ at the nanometre scale, based on graphene^{8–11}.

When a single graphite layer is terminated by zigzag edges on both sides, which we refer here to as a zigzag graphene nanoribbon (ZGNR) (Fig. 1), there are peculiar localized electronic states at each edge^{12,13}. These edge states (which are extended along the edge direction) decay exponentially into the centre of the ribbon, with decay rates depending on their momentum^{12–15}. Such states have been observed in monoatomic step edges of graphite by using scanning probe techniques^{16,17}. The localized edge states form a twofold degenerate flat band at the Fermi energy (E_F), existing in about one-third of the Brillouin zone away from the zone centre^{12–15}. By invoking band ferromagnetism, it has been suggested that an opposite spin orientation across the ribbon between ferromagnetically ordered edge states on each edge in ZGNRs is the ground-state spin configuration; that is, the total spin is zero^{12,18,19}. Because the states around E_F are the edge states and linear combinations of them, the effects of external transverse fields are expected to be significant on these states, in contrast with those on the extended states²⁰.

Our study of the spin-resolved electronic structure of ZGNRs is based on the *ab initio* pseudopotential density functional method²¹ within the local spin density approximation²². A periodic saw-tooth-type potential perpendicular to the direction of the ribbon edge is used to simulate the external electric fields (E_{ext}) in a supercell (Fig. 1). In accordance with previous convention^{12–15}, the ZGNRs are classified by the number of zigzag chains (n ; Fig. 1) forming the width of the ribbon. We will hereafter refer to an ZGNR with n zigzag chains as an n -ZGNR. When the spin degree of freedom is neglected, our calculation from first principles also predicts a twofold degenerate flat band at E_F (Fig. 2a). But the spinless state is not the ground state. Moreover, the electronic structures of the ZGNRs show marked alterations when spins and E_{ext} are included.

Considering first the spin degree of freedom, we find as in previous studies that the configuration with opposite spin (antiferromagnetic) orientation between ferromagnetically ordered edge states at each edge (Fig. 2b) is favoured as the ground state over the configuration with same spin orientation between the two edges^{12,18,19} (The present result of antiferromagnetic spin configuration on the honeycomb lattice is consistent with a theorem for electrons on a bipartite lattice²³.) Our calculations show that the magnetic interaction energies are quite large. For example, the total energy difference between a spin-polarized edge and a spin-unpolarized one is 20 meV per edge atom for 8-ZGNR, and the spin configuration is further stabilized by 2.0 meV per edge atom as a result of the antiferromagnetic coupling between the spin-polarized edges. Because the interaction between spins on opposite edges increases with decreasing width, the total energy of an n -ZGNR with antiferromagnetic arrangement across opposite edges is always lower than that of a ferromagnetic arrangement if $n \leq 32$. This total energy hierarchy is maintained when external electric fields are applied. It is known that spontaneous magnetic orderings in one-dimensional and two-dimensional spin lattice models are difficult to achieve at finite temperature²⁴. Spin correlation lengths comparable to nanometre-scale systems, however, are possible in practice^{25–28}. Here we also expect that the spin orderings are realizable because of the large anisotropic exchange interactions between the spins in ribbons with split-gate geometry on the substrate.

We find that the ground state of the ZGNRs, including the spin degree of freedom, has a bandgap inversely proportional to the ribbon width. However, the energy splitting at $ka = \pi$ is ~ 0.52 eV, regardless of width, if $n \geq 8$. The states of opposite spin orientation are degenerate in all bands (Fig. 2c, left). When spins are included, the

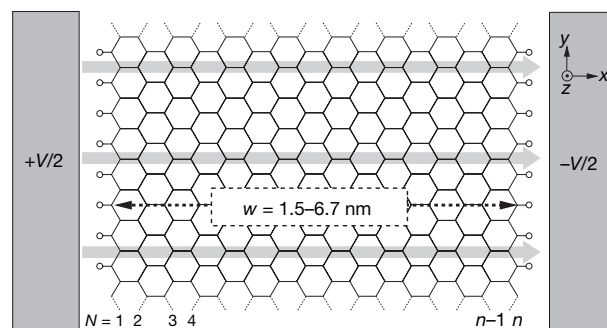


Figure 1 | Graphene nanoribbon in electric fields. Diagram of a zigzag graphene nanoribbon (ZGNR) with external transverse electric field E_{ext} . E_{ext} is applied across the ZGNR along the lateral direction (x direction) in an open-circuit split-gate configuration and is positive towards the right side. The ZGNRs are assumed to be infinite along the y direction. A small longitudinal source-drain field could be applied to generate spin-polarized currents along the y direction. Hydrogen atoms on the edges are denoted by circles. The ZGNR shown in this figure is a 16-ZGNR ($n = 16$). The width w of ZGNRs under study was in the range 1.5–6.7 nm.

¹Department of Physics, University of California at Berkeley, Berkeley, California 94720, USA. ²Materials Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA.

degeneracy between the occupied and unoccupied edge-state bands at E_F is now lifted and the edge states near E_F have dispersion along the direction of the edge with a bandwidth of ~ 2 eV when extended over the Brillouin zone.

With applied transverse electric fields, we find that the valence and conduction edge-state bands associated with one spin orientation close their gap, whereas those associated with the other widen theirs (Fig. 2c). So, under appropriate field strengths, the ZGNRs are forced into a half-metallic state by the applied electric field, resulting in insulating behaviour for one spin and metallic behaviour for the other. We shall defer the discussion of spin–orbit interactions later. For now, we label the gap-opening states as α -spin (shown in red in Figs 2–4) and the gap-narrowing states as β -spin (blue). In a 16-ZGNR, the bandgap associated with β -spin is completely closed by an E_{ext} of $0.1 \text{ V } \text{\AA}^{-1}$, whereas the gap for α -spin electrons remains very large at 0.30 eV (Fig. 2c). The energy gap for the β -spin electrons changes to an indirect gap from a direct gap as E_{ext} increases, and is closed indirectly (Fig. 2c, inset). After gap closure, an electron channel near $ka = \pi$ and a hole channel near $ka = 0.75\pi$ appear at E_F , all with the same spin direction.

The half-metallicity of the ZGNRs originates from the fact that the applied electric fields induce energy-level shifts of opposite signs for the spatially separated spin-ordered edge states. Such separate and opposite energy shifts are made possible by the localized nature of the edge states around E_F . Because oppositely oriented spin states are located at the opposite sides of the ZGNR, the effect of E_{ext} on them is opposite, moving the occupied and unoccupied β -spin states closer in energy but moving the occupied and unoccupied α -spin states apart (Fig. 3). The electrostatic potential is raised on the right side and lowered on the left side as $E_{\text{ext}} (>0)$ increases. Correspondingly, the energies for localized edge states on the right side are shifted upwards and those on the left side downwards, eventually leaving states of only one spin orientation at E_F (Fig. 3b). In a 16-ZGNR, the occupied α states and unoccupied β -spin states on the left side move downwards in energy by 19 meV and 110 meV, respectively, and occupied β -spin and unoccupied α -spin states on the right side upwards by 112 meV and 74 meV, respectively, as E_{ext} increases to

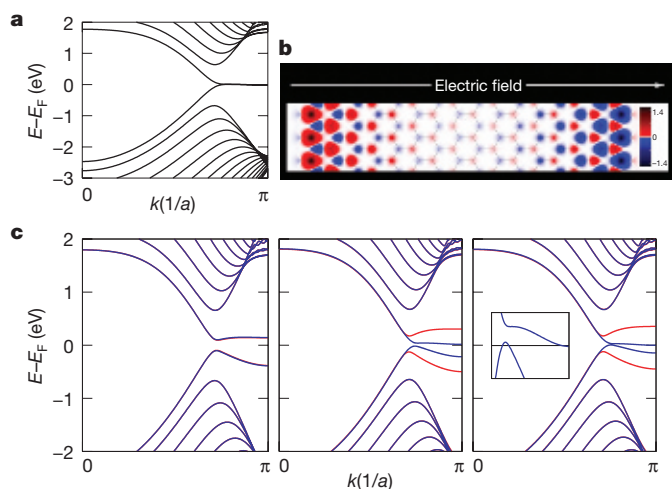


Figure 2 | Electronic structures of graphene nanoribbons. In all figures, the Fermi energy (E_F) is set to zero. **a**, The spin-unpolarized band structure of a 16-ZGNR. **b**, The spatial distribution of the charge difference between α -spin and β -spin ($\rho_\alpha(r) - \rho_\beta(r)$) for the ground state when there is no external field. The magnetization per edge atom for each spin on each sublattice is $0.43 \mu_B$ with opposite orientation, where μ_B is the Bohr magneton. The graph is the electron density integrated in the z direction, and the scale bar is in units of $10^{-2} |e| \text{ \AA}^{-2}$. **c**, From left to right, the spin-resolved band structures of a 16-ZGNR with $E_{\text{ext}} = 0.0, 0.05$ and $0.1 \text{ V } \text{\AA}^{-1}$, respectively. The red and blue lines denote bands of α -spin and β -spin states, respectively. Inset, the band structure with $E_{\text{ext}} = 0.1 \text{ V } \text{\AA}^{-1}$ in the range $|E - E_F| < 50 \text{ meV}$ and $0.7\pi \leq ka \leq \pi$ (the horizontal line is E_F).

$0.1 \text{ V } \text{\AA}^{-1}$ (Fig. 3c). The occupied β -spin states in the middle of a 16-ZGNR are the tails of the localized β -spin states on the right side and the unoccupied β -spin states are from the left side, so that occupied and unoccupied β -spin states in the middle of the ZGNR move oppositely to close the gap. The energies of the occupied and unoccupied α -spin states in the middle also follow movements of those of the corresponding localized states on each side, resulting in an increased gap value (Fig. 3c).

We note that the critical electric field for achieving half-metallicity in ZGNRs decreases as the width increases because the electrostatic potential difference between the two edges is proportional to the system size. For a 32-ZGNR whose width is $67.2 \text{ \AA}$, $E_{\text{ext}} = 0.045 \text{ V } \text{\AA}^{-1}$ is required to close the bandgap for the β -spin electrons (Fig. 4). Because the energy shifts of the edge states depend on the total voltage drop between the two sides, the variation of the energy gap is expected to exhibit a universal behaviour as a function of wE_{ext} , where w is the width of the ZGNR. This is seen in the inset in Fig. 4. From the calculations, the required critical field is estimated to

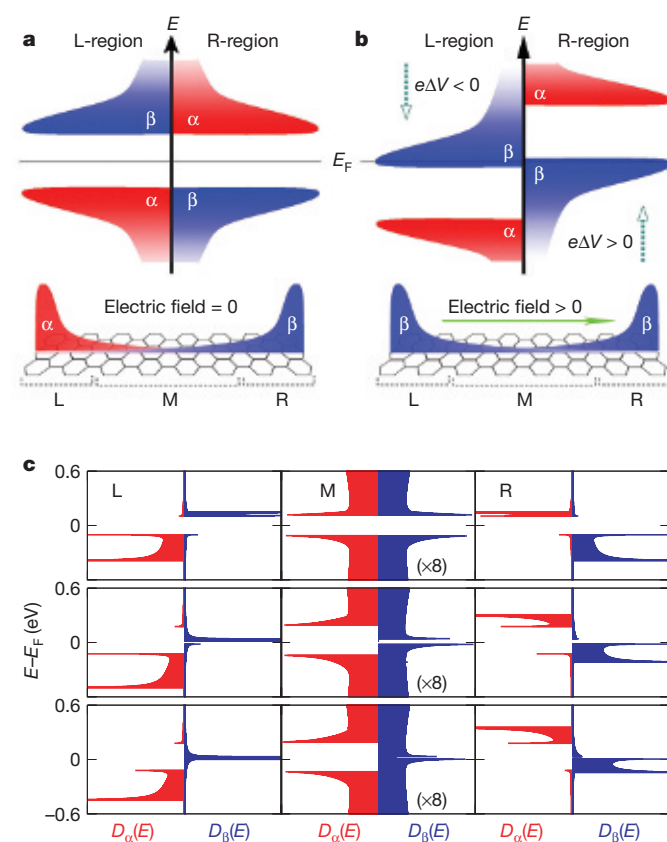


Figure 3 | Origin of half-metallicity. **a**, Schematic density-of-states diagram of the electronic states of a ZGNR in the absence of an applied electric field. Top: the occupied and unoccupied localized edge states on the left side (L-region as defined at the bottom) are the α -spin and β -spin states, respectively, and vice versa on the right side (R-region) with the same energy gap for both sides. Bottom: schematic diagram of the spatial spin distribution of the highest occupied valence band states without an external electric field. **b**, Top: with a transverse electric field, the electrostatic potential on the left edge is lowered ($e\Delta V < 0$), whereas the one on the right edge is raised ($e\Delta V > 0$). Correspondingly, the energies of the localized states at the left edge are decreased and those of the localized states at the right edge are increased. Bottom: the resulting states at E_F are only β -spin. **c**, From left to right, the local density of states of α -spins and β -spins (ordinate) of a 16-ZGNR as a function of energy (abscissa) for atoms in the L, M and R regions shown in **a**, respectively. From top to bottom, $E_{\text{ext}} = 0.0, 0.05$ and $0.1 \text{ V } \text{\AA}^{-1}$ respectively. The local density of states in the middle panels are enlarged eightfold for clarity. For $E_{\text{ext}} = 0.1 \text{ V } \text{\AA}^{-1}$, the van Hove singularities of the β -spin in the M and R region are above the E_F by 5 meV, and all states at E_F are of β -spin.

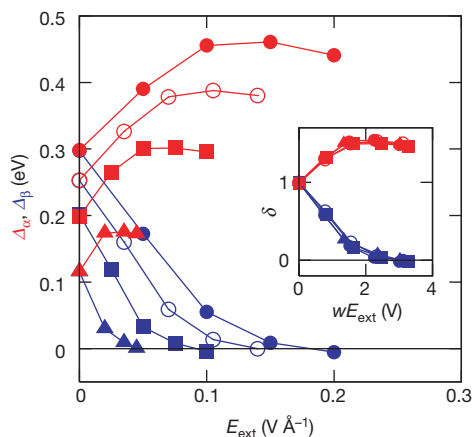


Figure 4 | Dependence of half-metallicity on system size. Δ_α (red) denotes the direct bandgap of α -spin, and Δ_β (blue) the (in)direct gap of β -spin as function of E_{ext} for the 8-ZGNR (filled circles), 11-ZGNR (open circles), 16-ZGNR (squares) and 32-ZGNR (triangles). The slope variation of Δ_α and Δ_β indicates the gap change from direct to indirect. The rescaled gaps $\delta_\alpha \equiv \Delta_\alpha(w, E_{\text{ext}})/\Delta_0(w)$ and $\delta_\beta \equiv \Delta_\beta(w, E_{\text{ext}})/\Delta_0(w)$ for the various widths collapse to a single function of wE_{ext} as shown in the inset, where $\Delta_\alpha(w, E_{\text{ext}})$ and $\Delta_\beta(w, E_{\text{ext}})$ are the bandgaps for the α -spins and β -spins, respectively, of the ZGNR with a width of w in $E_{\text{ext}} (\neq 0)$, and $\Delta_0(w) \equiv \Delta_\alpha(w, E_{\text{ext}} = 0) = \Delta_\beta(w, E_{\text{ext}} = 0)$.

be $3.0 \text{ (V)}/w \text{ (\AA)}$. To establish half-metallicity, the relevant energy scale is given by the field-induced energy shift, and its magnitude is in the order of 100 meV. Thus the small magnitude of spin-orbit interaction (4–6 meV) in carbon atoms^{29,30} would not change the half-metallic nature of the ZGNRs but would function in determining the spatial direction (normal direction with respect to the ribbon plane) of spin up and down in the ZGNRs²⁹.

Because edges are inevitably susceptible to defects, we have examined the robustness of the predicted half-metallicity to edge defects. Our calculations show that the system remains purely of one spin type at E_F in the presence of different types and concentrations of defects. Results on 8-ZGNRs with three different kinds of defect (dangling bonds, vacancies and Stone–Wales defects at 6–12% defect concentration per edge) are presented in Supplementary Fig. 1, confirming that the predicted half-metallicity is indeed robust.

Another consideration is that, when in the half-metallic state the ZGNRs are in a transverse electric field, the current-carrying electrons moving from the source to the drain in the longitudinal direction would experience an effective magnetic field due to spin–orbit interactions^{6,7} and the spins are expected to rotate. However, we find that the resulting extremely weak effective magnetic fields are parallel to the spatial spin direction (z direction in Fig. 1) already determined by the intrinsic spin–orbit interactions of carbon atoms. Suppose that we have the β -spin electrons moving with velocity $\mathbf{v} = v\hat{y}$ in $\mathbf{E} = E_{\text{ext}}\hat{x}$, the effective magnetic field exerted on the β -spin electrons would be $\mathbf{B}_{\text{eff}} = (ev\hbar/4mc^2)E_{\text{ext}}\hat{z}$, where $\hbar$ is the Planck constant, m is the mass of an electron, e is the charge on an electron, and c is the speed of light. At a critical electric field of $0.045 \text{ V \AA}^{-1}$ for a 32-ZGNR, the estimated energy for spin–orbit coupling due to E_{ext} is only $1.1 \times 10^{-4} \text{ meV}$. We also find that, as a result of the energy gap asymmetry for each spin, there is no spin precession even when the direction of E_{ext} is tilted or when the spatial spin direction is altered by spin–orbit interaction arising from the substrate. So, the spatial spin direction once determined would not change even if a strong transverse electric field were applied. This implies that, if we change the direction of E_{ext} , the spin polarity of the carriers at E_F of the half-metallic ribbon will be reversed because the induced potentials at the edges change their signs. Hence, under these conditions, the half-metallic nature is robust even though a transverse electric field is applied, and spin-polarized current should be obtained in a transport experiment with split gates.

Received 24 March 2006; accepted 16 August 2006.

- de Groot, R. A., Mueller, F. M., van Engen, P. G. & Buschow, K. H. J. New class of materials: half-metallic ferromagnets. *Phys. Rev. Lett.* **50**, 2024–2027 (1983).
- Park, J.-H. *et al.* Direct evidence for a half-metallic ferromagnet. *Nature* **392**, 794–796 (1998).
- Wolf, S. A. *et al.* Spintronics: A spin-based electronics vision for the future. *Science* **294**, 1488–1495 (2001).
- Fang, C. M., de Wijs, G. A. & de Groot, R. A. Spin-polarization in half-metals. *J. Appl. Phys.* **91**, 8340–8344 (2002).
- McEuen, P. L., Fuhrer, M. S. & Park, H. Single-walled carbon nanotube electronics. *IEEE Trans. Nanotechnol.* **1**, 78–85 (2002).
- Murakami, S., Nagaosa, N. & Zhang, S.-C. Dissipationless quantum spin current at room temperature. *Science* **301**, 1348–1351 (2003).
- Sinova, J. *et al.* Universal intrinsic spin Hall effect. *Phys. Rev. Lett.* **92**, 126603 (2004).
- Berger, C. *et al.* Ultrathin epitaxial graphite: 2D electron gas properties and a route toward graphene-based nanoelectronics. *J. Phys. Chem. B* **108**, 19912–19916 (2004).
- Novoselov, K. S. *et al.* Two-dimensional gas of massless Dirac fermions in graphene. *Nature* **438**, 197–200 (2005).
- Zhang, Y., Tan, Y.-W., Stormer, H. L. & Kim, P. Experimental observation of the quantum Hall effect and Berry's phase in graphene. *Nature* **438**, 201–204 (2005).
- Berger, C. *et al.* Electronic confinement and coherence in patterned epitaxial graphene. *Science* **312**, 1191–1196 (2006).
- Fujita, M., Wakabayashi, K., Nakada, K. & Kusakabe, K. Peculiar localized state at zigzag graphite edge. *J. Phys. Soc. Jpn.* **65**, 1920–1923 (1996).
- Nakada, K., Fujita, M., Dresselhaus, G. & Dresselhaus, M. S. Edge state in graphene ribbons: nanometer size effect and edge shape dependence. *Phys. Rev. B* **54**, 17954–17961 (1996).
- Wakabayashi, K., Fujita, M., Ajiki, H. & Sigrist, M. Electronic and magnetic properties of nanographite ribbons. *Phys. Rev. B* **59**, 8271–8282 (1999).
- Miyamoto, Y., Nakada, K. & Fujita, M. First-principles study of edge states of H-terminated graphitic ribbons. *Phys. Rev. B* **59**, 9858–9861 (1999).
- Kobayashi, Y., Fukui, K.-I., Enoki, T., Kusakabe, K. & Kaburagi, Y. Observation of zigzag and armchair edges of graphite using scanning tunneling microscopy and spectroscopy. *Phys. Rev. B* **71**, 193406 (2005).
- Niimi, Y. *et al.* Scanning tunneling microscopy and spectroscopy of the electronic local density of states of graphite surfaces near monoatomic step edges. *Phys. Rev. B* **73**, 085421 (2006).
- Okada, S. & Oshiyama, A. Magnetic ordering in hexagonally bonded sheets with first-row elements. *Phys. Rev. Lett.* **87**, 146803 (2001).
- Lee, H., Son, Y.-W., Park, N., Han, S. & Yu, J. Magnetic ordering at the edges of graphitic fragments: Magnetic tail interactions between the edge-localized states. *Phys. Rev. B* **72**, 174431 (2005).
- Son, Y.-W., Ihm, J., Cohen, M. L., Louie, S. G. & Choi, H. J. Electrical switching in metallic carbon nanotubes. *Phys. Rev. Lett.* **95**, 216602 (2005).
- Soler, J. M. *et al.* The SIESTA method for *ab initio* order- N materials simulation. *J. Phys. Condens. Matter* **14**, 2745–2779 (2002).
- Perdew, J. P. & Zunger, A. Self-interaction correction to density-functional approximations for many-electron systems. *Phys. Rev. B* **23**, 5048–5079 (1981).
- Lieb, E. H. Two theorems on the Hubbard model. *Phys. Rev. Lett.* **62**, 1201–1204 (1989).
- Mermin, N. D. & Wagner, H. Absence of ferromagnetism or antiferromagnetism in one- or two-dimensional isotropic Heisenberg models. *Phys. Rev. Lett.* **17**, 1133–1136 (1966).
- Gambardella, P. *et al.* Ferromagnetism in one-dimensional monatomic metal chains. *Nature* **416**, 301–304 (2002).
- Dorantes-Dávila, J. & Pastor, G. M. Magnetic anisotropy of one-dimensional nanostructures of transition metals. *Phys. Rev. Lett.* **81**, 208–211 (1998).
- Vindigni, A., Rettori, A., Pini, M. G., Carbone, C. & Gambardella, P. Finite-sized Heisenberg chains and magnetism of one-dimensional metal systems. *Appl. Phys. A* **82**, 385–394 (2006).
- Delin, A., Tossati, E. & Weht, R. Magnetism in atomic-size palladium contacts and nanowires. *Phys. Rev. Lett.* **92**, 057201 (2004).
- Yao, Y., Ye, F., Qi, X.-L., Zhang, S.-C. & Fang, Z. Spin-orbit gap of graphene. Preprint at <http://arxiv.org/abs/cond-mat/0606350> (2006).
- Min, H. *et al.* Intrinsic and Rashba spin-orbit interactions in graphene sheets. Preprint at <http://arxiv.org/abs/cond-mat/0606504> (2006).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank J. Neaton, F. Giustino, I. Souza, C. H. Park and H. J. Choi for discussions. This research was supported by the National Science Foundation (NSF) and by the Director, Office of Science, Office of Basic Energy Science, Division of Material Sciences and Engineering, US Department of Energy (DOE). Computational resources have been provided by the NSF at the National Partnership for Advanced Computational Infrastructure and by the DOE at the National Energy Research Scientific Computing Center.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to S.G.L. (sglouie@berkeley.edu).

LETTERS

Free-electron-like dispersion in an organic monolayer film on a metal substrate

R. Temirov¹, S. Soubatch¹, A. Luican¹ & F. S. Tautz¹

Thin films of molecular organic semiconductors are attracting much interest for use in electronic and optoelectronic applications. The electronic properties of these materials and their interfaces are therefore worth investigating intensively^{1–3}, particularly the degree of electron delocalization that can be achieved^{2,4}. If the delocalization is appreciable, it should be accompanied by an observable electronic band dispersion. But so far only limited experimental data on the intermolecular dispersion of electronic states in molecular materials is available^{5–8}, and the mechanism(s) of electron delocalization in molecular materials are also not well understood. Here we report scanning tunnelling spectroscopy observations of an organic monolayer film on a silver substrate, revealing a completely delocalized two-dimensional band state that is characterized by a metal-like parabolic dispersion with an effective mass of $m^* = 0.47m_e$, where m_e is the bare electron mass. This dispersion is far stronger than expected for the organic film alone⁷, and arises as a result of strong substrate-mediated coupling between the molecules within the monolayer.

Our experiments have been carried out on monolayer films of PTCDA on Ag(111). PTCDA, or 3,4,9,10-perylenetetracarboxylic-acid-dianhydride, is an archetypal organic semiconductor^{9–11} (see Fig. 1a for the structure of the molecule and its bulk crystal structure¹²) and one of the few materials of this type for which band dispersion data have been obtained⁷. Along the π -stacking direction—the direction of maximum dispersion in the bulk—a moderate dispersion with a perpendicular effective mass of $m_{\perp}^* = 5.28m_e$ (see Fig. 1a) has been found (bandwidth 200 meV)⁷; any intrinsic dispersion within the layers of molecules (see Fig. 1a) has so far proved too small for detection (bandwidth ≤ 100 meV). The high degree of order at the epitaxial PTCDA/Ag(111) interface⁹ (see Methods and Supplementary Information a) makes this system well-suited for a dispersion study of a molecular monolayer on a metal.

For the dispersion measurements we have used scanning tunnelling spectroscopy (STS). Although the wave vectors of electrons tunnelling between tip and sample are not directly accessible, the atomic-scale spatial resolution of the scanning tunnelling microscope (STM) can be converted into a wave vector resolution to determine dispersions of itinerant electron states experimentally. The trick is to prepare a molecular film consisting of nanometre-sized islands. The quantum mechanical electron confinement in these islands selects discrete wave vectors for which the itinerant state can form a standing wave within an island of given size and shape¹³. Local spectroscopy on these islands then reveals the electronic state energies for the selected k .

STS spectra detailing the local electronic structure of the PTCDA monolayer film on Ag(111) are presented in Fig. 1b. The feature P1 arises from the lowest unoccupied molecular orbital (LUMO) of PTCDA. On adsorption, hybridization with metal states causes

substantial broadening of the LUMO, relative to that of the free molecule. At the same time, the LUMO is shifted below the Fermi energy and thus becomes partially occupied¹⁴. This behaviour follows the classical theory of weak chemisorption developed originally for small molecular adsorbates¹⁵. The differential LUMO shift and its partial occupation upon PTCDA adsorption have also been found in a range of independent spectroscopic measurements^{16–18}.

The unoccupied spectral feature P2 is derived from the LUMO+1 of the free molecule. Its distinctive shape, that is, its very steep initial rise and nearly constant level from then on, is reminiscent of the theoretical density of states (DOS) for a two-dimensional free electron band^{19,20}. To verify this interpretation, and to determine P2's dispersion quantitatively, we make use of the electron confinement effect described above in an ensemble of nanometre-sized monolayer islands of PTCDA on Ag(111), such as is shown in Fig. 2a (see Methods for their preparation).

Figure 2c shows a close-up of an island consisting of 39 molecules in the familiar herringbone arrangement. Spectra taken at two different spots inside this island (Fig. 2b) exhibit a rather discrete peak structure instead of the continuous band seen in Fig. 1b, which is exactly as expected if the itinerant electron state P2 is confined to a small finite island. This is confirmed by spectroscopic images (Fig. 2d–f) recorded at appropriate bias voltages. They clearly show the $n = 1$, $n = 2$, and $n = 3$ resonances of a confined electron state (Supplementary Fig. 3 shows a simulation of confined states for qualitative comparison with Fig. 2d–f).

Ensembles of islands (for example, Fig. 2a) allow us to analyse the peak positions of the $n = 1, 2, \dots$ states as a function of island sizes. As shown by the results in Fig. 2g, we find that the quantization energy varies linearly with inverse island size as predicted by confinement theory (see Methods). When plotting the dispersion $E(k)$ directly, we

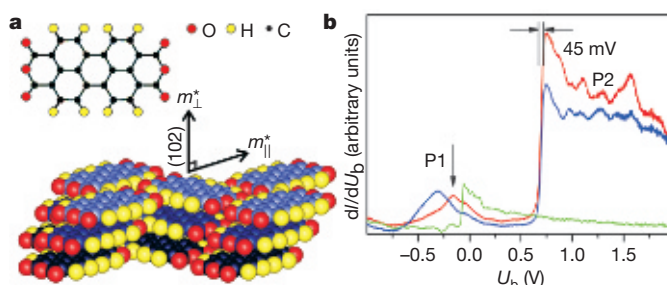


Figure 1 | Structure and electronic properties of PTCDA/Ag(111). **a**, Bulk crystal structure (α -phase)¹². Within the (102)-planes, molecules are packed in a herringbone motif. **b**, STS spectra ($V_{\text{mod}} = 4$ mV, $\nu_{\text{mod}} = 1,233$ Hz, $U_b = -340$ mV, $I = 0.16$ nA) of a PTCDA/Ag(111) monolayer. Red and blue spectra were measured on inequivalent type A and type B molecules in the superstructure cell; see Supplementary Information section a. Green line is the STS spectrum of the Ag surface state.

¹School of Engineering and Science, International University Bremen, P. O. Box 750561, 28725 Bremen, Germany.

obtain the free-electron dispersion parabola shown in Fig. 2h (see Methods). Fitting the data in Fig. 2g and h allows us to determine an effective mass of $m_{||}^* = 0.47m_e$, which corresponds to a bandwidth of at least 270 meV (Fig. 2h). Remarkably, these values indicate that the dispersion of the delocalized band P2—that is, the dispersion within the plane of a monolayer of supposedly weakly interacting molecules (Fig. 1a)—is more than ten times larger than the maximum dispersion in bulk PTCDA (characterized by $m_{\perp}^*$)⁷.

We note that the effective mass $m_{||}^* = 0.47m_e$ of the P2 state is similar to the effective mass $m^* = 0.42m_e$ of the Ag(111) surface state (for example, ref. 13), raising questions about the relationship between these states in our system. First, we find that the latter is not only depopulated, but also strongly scattered and destroyed by PTCDA adsorbates, as seen in spectroscopic images of the Ag(111) surface state close to PTCDA island boundaries and single molecules (data not shown). Moreover, spectroscopic mapping of the spatial distributions of the P1 and P2 states reveals that P2 is concentrated on the molecules (Fig. 3a; see also Methods for details). We thus conclude that outside the surface Bloch states formed by molecular orbitals with a spatial distribution that resembles that of LUMO+1 of PTCDA are contributing to the wavefunction of P2. Further experimental evidence for a significant molecular contribution to P2 is provided by spectroscopic mapping of P2 close to an island boundary (Supplementary Information b). The silver surface is of course expected to participate in the PTCDA/Ag(111) interface state P2; but the finding that disordered PTCDA layers show an entirely dif-

ferent spectroscopic signature (L. Kilian *et al.*, manuscript in preparation) indicates that an ordered array of molecular LUMO+1 orbitals is essential for the formation of P2 in its presently observed form.

Our data show that the properties of P2 depend sensitively on molecular coordination (and hence on coupling between the constituent molecular orbitals) within the organic monolayer. The effect is illustrated for an island consisting of three domains, two with a herringbone structure and one with a square structure (Fig. 4a; see also Methods): the STS spectra and images (Fig. 4b–d) show that the confined states of the square domain occur at energies different from those of the corresponding confined states of the herringbone domains. A quantitative analysis yields a dispersion parabola for the square phase that is shifted from the parabola for the herringbone phase, by 80 meV to lower energies (Fig. 2g and h). The square phase possibly has a slightly larger effective mass than the herringbone phase, but this cannot be established unambiguously because square-phase islands occur in a limited size range only.

The intermolecular coupling giving rise to the remarkable dispersion seen in our data could be either a direct coupling between molecular orbitals, or an indirect coupling mediated by the substrate. It is quite likely that the compression of the PTCDA monolayer enforced by substrate commensurability⁹ (see Supplementary Information b) will enhance direct interaction between neighbouring LUMO+1 orbitals; but it seems unlikely that this effect is sufficient to explain the strong dispersion we see in this system. In fact, substrate-mediated coupling (schematically shown in Fig. 3b by green dotted lines) is strongly suggested by a coincidence of two experimental observations. First, STS imaging (Fig. 3a, b) shows that unlike the molecular orbitals contributing to P1, the P2-forming LUMO+1 additionally involves the anhydride functional groups of PTCDA.

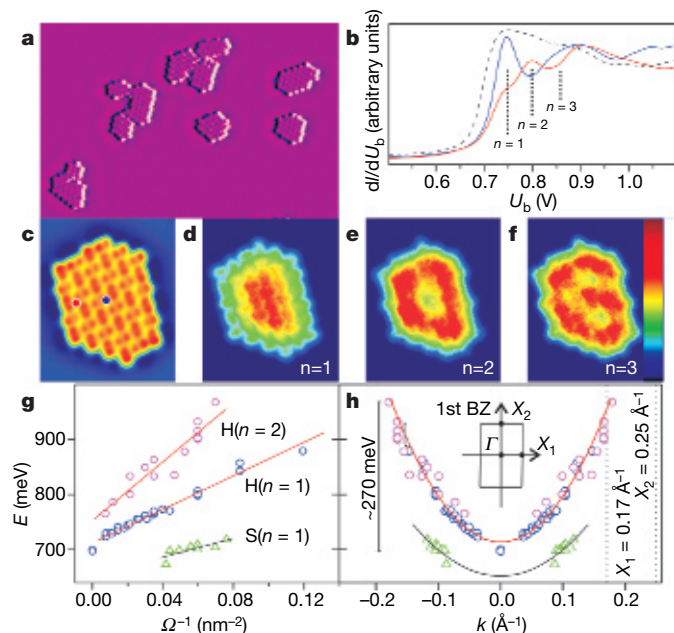


Figure 2 | Free-electron-like two-dimensional band state of PTCDA/Ag(111). **a**, STM image ($64 \times 42 \text{ nm}^2$, $I = 30 \text{ pA}$, $U_b = -340 \text{ mV}$) of small ordered PTCDA islands on Ag(111). **b**, Solid lines are STS spectra ($V_{\text{mod}} = 4 \text{ mV}$, $v_{\text{mod}} = 1,733 \text{ Hz}$, $U_b = -340 \text{ mV}$, $I = 0.2 \text{ nA}$, ten-point-smoothed) of the PTCDA island presented in **c**, with vertical dotted lines indicating its first three confined states. Dashed line is the monolayer spectrum from Fig. 1b. **c**, STM image ($10 \times 12.5 \text{ nm}^2$, $I = 0.2 \text{ nA}$, $U_b = -340 \text{ mV}$) of a 39-molecule herringbone island. **d–f**, Corresponding STS images ($V_{\text{mod}} = 4 \text{ mV}$, $v_{\text{mod}} = 1,733 \text{ Hz}$, $I = 0.2 \text{ nA}$), recorded at voltages 750 mV ($n = 1$), 800 mV ($n = 2$) and 860 mV ($n = 3$), with the colour-coding used throughout this Letter. **g**, Energies of the $n = 1$ (herringbone, blue) and $n = 2$ (herringbone, red) confined states versus inverse island area. Green indicates the $n = 1$ confined state of square phase islands. **h**, Parabolic dispersion (mirrored at $k = 0$) of the herringbone and square phase layers. The black vertical error bar indicates the variation of the dispersion parabola if hexagonal or square islands are used for converting Ω to k instead of circular ones. Inset, the first Brillouin zone (BZ) of PTCDA superstructure.

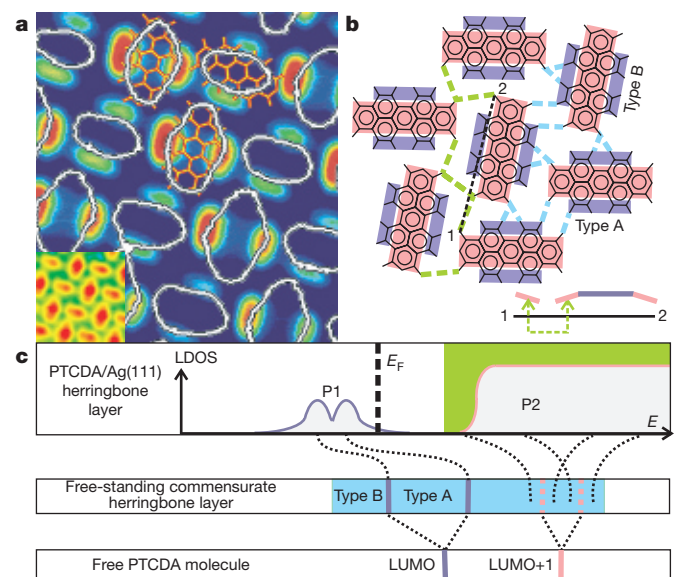


Figure 3 | Wavefunctions and energy level scheme of P1 and P2. **a**, The coloured spectroscopic image in the background has been recorded at the peak P1 (black arrow in Fig. 1b), while the white outline corresponds to a spectroscopic image recorded close to the onset energy of P2 (the full image is shown in the inset). Background image: $4.5 \times 4.5 \text{ nm}^2$, $V_{\text{mod}} = 10 \text{ mV}$, $v_{\text{mod}} = 1,233 \text{ Hz}$, $I = 0.5 \text{ nA}$, $U_b = -160 \text{ mV}$. Inset, same parameters as background image except $U_b = 730 \text{ mV}$. White outlines are contours (50%-lines) of the molecules in the inset image projected into the main image. All images have been divided by the barrier transmission function (see Methods). **b**, Schematic representation of the images in **a**. Pink and purple shadings schematically indicate areas in which the LUMO (purple) and LUMO+1 (pink) of the free molecule are concentrated. Green lines indicate substrate-mediated interaction, leading to dispersion of P2. Blue lines indicate Davydov interactions, leading to splitting of P1. **c**, Schematic energy level scheme (colour-coded).

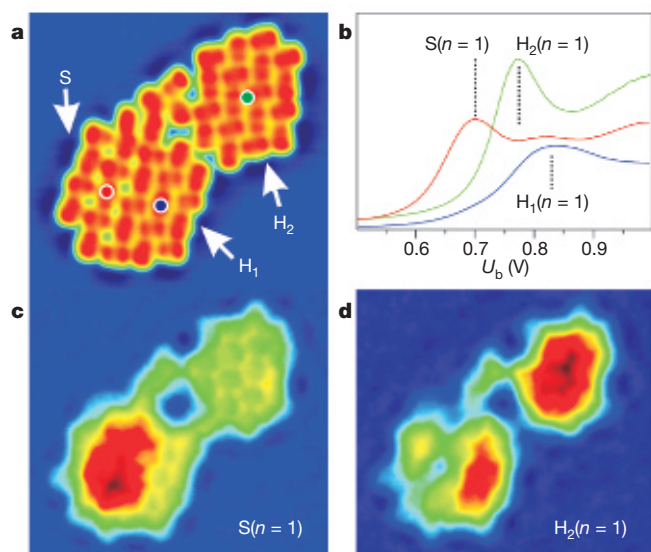


Figure 4 | Band state in herringbone and square phase islands. **a**, STM image ($14 \times 13 \text{ nm}^2$, $I = 0.2 \text{ nA}$, $U_b = -340 \text{ mV}$) of a PTCDA island on Ag(111) containing domains of different structural order: H, herringbone phase; S, square phase. **b**, STS spectra ($V_{\text{mod}} = 4 \text{ mV}$, $\nu_{\text{mod}} = 1,733 \text{ Hz}$, $U_b = -0.340 \text{ mV}$, $I = 0.2 \text{ nA}$, ten-point-smoothed) of the square phase domain (red) and the two herringbone domains (green, blue), recorded at positions marked in **a**. **c–d**, Corresponding STS images ($14 \times 13 \text{ nm}^2$, $V_{\text{mod}} = 4 \text{ mV}$, $\nu_{\text{mod}} = 1,733 \text{ Hz}$, $I = 0.2 \text{ nA}$), recorded at $U_b = 700 \text{ mV}$ and 780 mV , respectively.

Second, it is known that in monolayers on Ag(111), PTCDA is internally distorted so that the centre of the molecule is further away from the substrate than the carboxylic oxygen atoms^{21–23} (see scheme at bottom of Fig. 3b), and moreover the adsorption height of the corner oxygen atoms is susceptible to the molecular coordination (L. Kilian *et al.*, manuscript in preparation). In combination with the sensitivity of the P2 band to the intermolecular coordination, these observations imply that the intermolecular coupling of the band-forming LUMO+1 orbitals is critically dependent on the interaction of the molecules with the substrate, as expected for a substrate-mediated interaction mechanism.

In contrast to P2, P1 does not exhibit a measurable dispersion here; this state, which is responsible for chemical adsorption of the molecule to the metal substrate^{14,16,17}, is thus rather localized. But intermolecular couplings still influence P1, albeit differently from P2: the commensurability of the molecular superstructure with the substrate enforces the existence of two non-equivalent adsorption sites and hence molecules of type A and B within the monolayer (Supplementary Information a), which results in Davydov splitting of the LUMO levels for type A and type B molecules¹⁴. Some of the corresponding interactions are indicated by blue dotted lines in Fig. 3b. A schematic energy level scheme, comparing the different behaviour of P1 and P2 and their genesis from the respective molecular levels (LUMO, LUMO+1) is displayed in Fig. 3c.

Taken together, our observations demonstrate that organic–metal hybrid materials, modelled here by a single organic–metal interface, may provide a way of influencing intermolecular interactions, by taking advantage of the modified electronic structure of the interface region. This has been suggested as a general concept for enhancing electron dispersion (and thus electron delocalization) in organic materials¹.

METHODS

Thin film deposition. PTCDA was deposited by organic molecular beam deposition from home-built Knudsen cells heated to 430°C with a deposition rate of 0.5 monolayers per minute onto Ag(111) surfaces held at room temperature. Deposition was followed by an annealing step at 550 K . This protocol leads to the

formation of very large, almost perfectly ordered herringbone islands⁹. If PTCDA is deposited on Ag(111) surfaces held at low temperature (100 K), a disordered phase forms in which the molecules are arranged in random clusters. This phase has distinct electronic and structural properties which will be discussed elsewhere (L. Kilian *et al.*, manuscript in preparation). The small islands shown in Fig. 2a have been prepared by first depositing a disordered layer with $\sim 10\%$ coverage and then annealing the surface at 190 K for 2 min. After this procedure, most islands show a clear herringbone pattern, while a small fraction of the islands exhibits a square arrangement of molecules as an intermediate metastable phase. In the square phase the local packing geometry of the molecules differs from the herringbone phase (Fig. 4a).

STS spectroscopy. All STM/STS have been recorded at $6\text{--}9 \text{ K}$ in a Createc low-temperature STM. STS spectroscopy is carried out at constant tip height with the feedback loop switched off and lock-in detection of the differential conductivity dI/dU . For each spectrum, the parameters modulation amplitude (V_{mod}), modulation frequency (ν_{mod}), bias voltage (U_b) and tunnelling current (I) at which the tip was placed above the surface before opening the feedback loop are given in the figure legends. During STS imaging the differential conductivity dI/dU is measured at each image pixel of a scan performed at a specific tunnelling bias in constant current mode. This excludes vertical drift. The changes in tip height from image pixel to pixel of the spectroscopic image lead to a superposition of the spectroscopic image with a ‘topographic’ constant current image in the raw data. The latter has been divided out of raw images corresponding to the spectroscopic images presented in Fig. 2d–f. The images in Fig. 3 have been divided by the appropriate tunnelling barrier transmission function determined experimentally to be proportional to $e^{-2\kappa z}$ with $2\kappa = 1.7 \text{ \AA}^{-1}$.

Dispersion and effective mass. The dispersion is determined by plotting quantization energies $E_n - E_0$ for $n = 1, 2, \dots$ states versus inverse island size Ω^{-1} and fitting with the following equation: $E_n - E_0 = \frac{\hbar^2 \lambda_n^2}{m^* \Omega}$ (ref. 13). The island size Ω is determined by multiplying the number of molecules with the area per molecule in the herringbone structure $\sim 1.195 \text{ nm}^2$. Data for islands in the size range 7 to 108 molecules have been analysed. For the square phase the area per molecule has been determined from STM images as $\sim 1.348 \text{ nm}^2$. The parameters λ_n are shape parameters. For circular islands $\lambda_1 = 9.04$, and $\lambda_2 = 22.95$ (ref. 24), while for square islands $\lambda_1 = 9.87$ and for hexagonal islands $\lambda_1 = 9.30$ (ref. 13). Because our PTCDA islands on Ag(111) have irregular shapes, we have used the circular λ_n for all islands. The resulting error in m_1^* is $< 10\%$, because the differences between the λ_n for different shapes are small. The linear plot of Fig. 2g has been converted to a dispersion parabola $E_n(k)$ by applying the following equation: $k^2 = \frac{2\lambda_n}{\Omega}$.

Received 26 April; accepted 18 September 2006.

- Dimitrakopoulos, C. D. & Malenfant, P. R. L. Organic thin film transistors for large area electronics. *Adv. Mater.* **14**, 99–117 (2002).
- Karl, N. Charge carrier transport in organic semiconductors. *Synth. Met.* **133–134**, 649–657 (2003).
- Cahen, D., Kahn, A. & Umbach, E. Energetics of molecular interfaces. *Mater. Today* **8**, 32–41 (2005).
- Sun, Y., Yunqi, L. & Daoben, Z. Advances in organic field-effect transistors. *J. Mater. Chem.* **15**, 53–65 (2005).
- Gensterblum, G. *et al.* Experimental evidence for 400-meV valence band dispersion in solid C_{60} . *Phys. Rev. B* **48**, 14756–14759 (1993).
- Hasegawa, S. *et al.* Intermolecular energy-band dispersion in oriented thin films of bis(1,2,5-thiadiazolo)-p-quinobis(1,3-dithiole) by angle-resolved photo-emission. *J. Chem. Phys.* **100**, 6969–6973 (1994).
- Yamane, H. *et al.* Intermolecular energy-band dispersion in PTCDA multilayers. *Phys. Rev. B* **68**, 033102(4) (2003).
- Koch, N. *et al.* Evidence for temperature-dependent electron band dispersion in pentacene. *Phys. Rev. Lett.* **96**, 156803(4) (2006).
- Glöckler, K. *et al.* Highly ordered structures and submolecular scanning tunneling microscopy contrast of PTCDA and DM-PBDCI monolayers on Ag(111) and Ag(110). *Surf. Sci.* **405**, 1–20 (1998).
- Witte, G. & Wöll, C. Growth of aromatic molecules on solid substrates for applications in organic electronics. *J. Mater. Res.* **19**, 1889–1916 (2004).
- Forrest, S. R. Ultrathin organic films grown by organic molecular beam deposition and related techniques. *Chem. Rev.* **97**, 1793–1896 (1997).
- Möbus, M., Karl, N. & Kobayashi, T. Structure of perylene-tetracarboxylic-dianhydride thin films on alkali halide crystal substrates. *J. Cryst. Growth* **116**, 495–504 (1992).
- Li, J. T., Schneider, W. D., Berndt, R. & Crampin, S. Electron confinement to nanoscale Ag islands on Ag(111): A quantitative study. *Phys. Rev. Lett.* **80**, 3332–3335 (1998).
- Kraft, A. *et al.* Lateral adsorption geometry and site-specific electronic structure of a large organic chemisorbate on a metal surface. *Phys. Rev. B* **74**, 041402(R) (2006).

15. Norskov, J. K. Chemisorption on metal surfaces. *Rep. Prog. Phys.* **53**, 1253–1295 (1990).
 16. Jung, M. *et al.* The electronic structure of adsorbed aromatic molecules: Perylene and PTCDA on Si(111) and Ag(111). *J. Molec. Struct.* **293**, 239–244 (1993).
 17. Zou, Y. *et al.* Chemical bonding of PTCDA on Ag surfaces and the formation of interface states. *Surf. Sci.* **600**, 1240–1251 (2006).
 18. Eremitchenko, M., Schaefer, J. A. & Tautz, F. S. Understanding and tuning the epitaxy of large aromatic adsorbates by molecular design. *Nature* **425**, 602–605 (2003).
 19. Ando, T., Fowler, A. B. & Stern, F. Electronic properties of two-dimensional systems. *Rev. Mod. Phys.* **54**, 437–672 (1982).
 20. Kröger, J. *et al.* Surface state electron dynamics of clean and adsorbate-covered metal surfaces studied with the scanning tunnelling microscope. *Prog. Surf. Sci.* **80**, 26–48 (2005).
 21. Hauschild, A. *et al.* Molecular distortions and chemical bonding of a large π -conjugated molecule on a metal surface. *Phys. Rev. Lett.* **94**, 036106(4) (2005).
 22. Rurali, R., Lorente, N. & Ordejón, P. Comment on 'Molecular distortions and chemical bonding of a large π -conjugated molecule on a metal surface'. *Phys. Rev. Lett.* **95**, 209601 (2005).
 23. Hauschild, A. *et al.* Reply to "Comment on 'Molecular distortions and chemical bonding of a large π -conjugated molecule on a metal surface'". *Phys. Rev. Lett.* **95**, 209602 (2005).
 24. Kaufman, D. L., Kosztin, I. & Schulten, K. Expansion method for stationary states of quantum billiards. *Am. J. Phys.* **67**, 133–141 (1999).
- Supplementary Information** is linked to the online version of the paper at www.nature.com/nature.
- Acknowledgements** The work was financially supported by the Deutsche Forschungsgemeinschaft.
- Author Contributions** R.T. and S.S. conducted the experiments and prepared the figures. A.L., in the context of her B.Sc. thesis, participated in the analysis of the data on the delocalized state. F.S.T. and R.T. wrote the paper. R.T., S.S. and F.S.T. discussed the experiments, the data analysis, and the manuscript intensively at all stages of the work.
- Author Information** Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to F.S.T. (s.tautz@iu-bremen.de).

LETTERS

Flushing submarine canyons

Miquel Canals¹, Pere Puig², Xavier Durrieu de Madron³, Serge Heussner³, Albert Palanques² & Joan Fabres^{1†}

The continental slope is a steep, narrow fringe separating the coastal zone from the deep ocean. During low sea-level stands, slides and dense, sediment-laden flows erode the outer continental shelf and the continental slope, leading to the formation of submarine canyons that funnel large volumes of sediment and organic matter from shallow regions to the deep ocean¹. During high sea-level stands, such as at present, these canyons still experience occasional sediment gravity flows^{2–5}, which are usually thought to be triggered by sediment failure or river flooding. Here we present observations from a submarine canyon on the Gulf of Lions margin, in the northwest Mediterranean Sea, that demonstrate that these flows can also be triggered by dense shelf water cascading (DSWC)—a type of current that is driven solely by sea-water density contrast. Our results show that DSWC can transport large amounts of water and sediment, reshape submarine canyon floors and rapidly affect the deep-sea environment. This cascading is seasonal, resulting from the formation of dense water by cooling and/or evaporation, and occurs on both high- and low-latitude continental margins^{6–8}. DSWC may therefore transport large amounts of sediment and organic matter to the deep ocean. Furthermore, changes in the frequency and intensity of DSWC driven by future climate change may have a significant impact on the supply of organic matter to deep-sea ecosystems and on the amount of carbon stored on continental margins and in ocean basins.

An intricate network of submarine canyons with heads cut in the 130-m-deep crescent-shaped shelf is the most outstanding seafloor feature of the Gulf of Lions. Water is transported in a cyclonic direction by a thermo-haline along-slope current and a wind-driven mean coastal circulation. Constrained by the slope current offshore and the coast inshore, most shelf water is funnelled towards the narrowing southwestern shelf end where it hits the Cap de Creus promontory and is thereby deviated towards the nearby canyon (Fig. 1a).

Winter heat losses and evaporation induced by cold and dry northerly winds cause cooling and mixing of the Gulf of Lions' coastal and off-shelf waters. Once denser than surrounding waters, shelf water sinks, overflows the shelf edge, and cascades downslope until it reaches its equilibrium depth. Winter DSWC appears to be a major export mechanism with a strong inter-annual variability (Supplementary Fig. 1a). Cascading rapidly advects dense shelf water hundreds of metres deep over the slope where it merges with dense water formed off-shelf⁹. Since the early 1950s, winter hydrological surveys traced shelf water tongues within Lacaze-Duthiers canyon (LDC), next to Cap de Creus canyon (CCC)¹⁰, with equilibrium depths between 170 and 800 m. Continuous monitoring of temperature and current since 1993 in LDC showed that shelf water sunk down to 500 m almost every winter¹⁰. During the 1998–99 and 2004–05 abnormally cold and windy winters, further characterized by lower ($\sim 22 \text{ km}^3$) than average ($\sim 30 \text{ km}^3$) northern freshwater inputs

limiting the gain of buoyancy, it passed 1,000 m and was associated with exceptionally large sediment transfer (Supplementary Fig. 1).

In winter 2003–04, when seven canyon heads in the Gulf of Lions were monitored simultaneously (Fig. 1), the down-canyon cumulative sediment transport in CCC (with fluxes up to 3 t m^{-2} for the 3-day-long strongest flushing outburst in late February) was one to two orders of magnitude higher than in all other canyons (Supplementary Fig. 2). We therefore focused our winter 2004–05 observations on this canyon (Fig. 1b). A major DSWC episode occurred from late February to late March 2005. DSWC outbursts were characterized by significant temperature decreases, and by

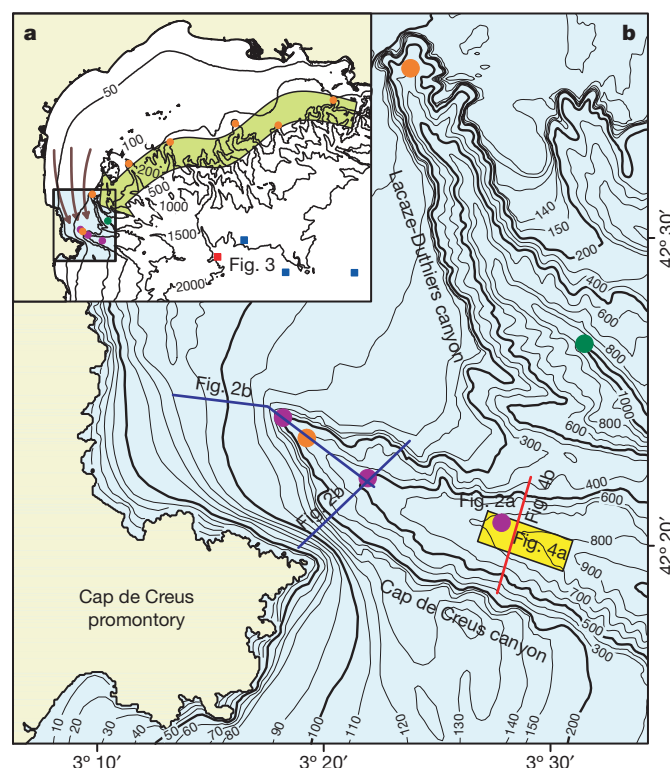


Figure 1 | Bathymetry maps and station location. **a**, Bathymetry map of the Gulf of Lions (GL). Mooring stations shown as follows. Orange dots, winter 2003–04 at 300 m; purple, winter 2004–05 at 200, 500 and 750 m; and green, long term mooring at 1,000 m. Deep hydrological stations at (red square) and off (blue squares) the mouth of the Cap de Creus Canyon (CCC) are also shown (see Fig. 3). Arrows indicate the direction of the mean coastal (brown) and slope circulation (green). **b**, Detailed bathymetry of the southwestern end of the GL. Time series shown in Fig. 2a correspond to the 750 m mooring (purple dot) inside CCC. Locations of the hydrological sections in Fig. 2b (blue lines), and of the side scan sonar sonograph (yellow box) and section b (red line) in Fig. 4 are also indicated.

¹CRG Marine Geosciences, Department of Stratigraphy, Paleontology and Marine Geosciences, University of Barcelona, E-08028 Barcelona, Spain. ²Marine Sciences Institute, CSIC, E-08003 Barcelona, Spain. ³CEFREM, UMR 5110 CNRS-University of Perpignan, F-66860 Perpignan Cedex, France. †Present address: Marine Sciences Research Center, Stony Brook University, Stony Brook, New York 11794-5000, USA.

increases in down-canyon current speed, water density and suspended sediment concentration (SSC) (Fig. 2a and Supplementary Fig. 3). Current direction and hydrological characteristics during the cascading event highlighted the asymmetric path of the dense and turbid water tongue sweeping the southern canyon wall (Fig. 2b). While a new survey in April 2005 confirmed its disappearance from the upper and middle canyon reaches, a 2,141 m deep cast at the mouth of the canyon revealed the near bottom intrusion of colder, fresher and turbid dense shelf water below the deep water newly formed off-shelf (Fig. 3). The magnitude of the winter 2004–05 DSWC was comparable to the 1998–99 event and had a strong impact on the deepest water masses^{11,12}.

Silt and sand-sized bed loads associated with such extreme events erode canyon floors. We found a field of giant furrows—tens of kilometres long, with wavelengths up to 100 m and heights up to 10 m (Fig. 4)—covering most of the CCC floor down to 1,400 m. These megascale bedforms, carved on overconsolidated mud, are organized into sets with different directions and degrees of development. The field hangs >50 m over a >500-m-wide sand and gravel-filled axial incision ('thalweg') collecting particles transported along the furrows and the canyon axis upstream. Its location corresponds to the area directly affected by DSWC during the severe 2005 event, including parts of the southern canyon wall (Fig. 2b). *In situ* measurements, sonographic evidence (Fig. 4), and published criteria^{13,14} confirm that erosion prevails in the furrowed area and that the furrow formation process is currently active, though intermittent.

Giant furrows are erosive features requiring highly energetic processes to develop^{13,14}. The current shear stress generated during the 2005 cascading event was large enough ($\sim 0.7 \text{ N m}^{-2}$) to resuspend sand, which itself has the potential to erode fine-grained cohesive substrata to form furrows^{13,14}. During the main cascading event, the mean grain size of the sediment caught by traps deployed 30 m above the bottom in the canyon axis ranged from 31 to 62 μm with >50% silt and sand. Sediment collected before cascading had a size of 2–4 μm with <1% silt and sand (Supplementary Fig. 4). Both sediments were, however, finer than those in the axial incision (0.5–8 mm coarse sand to gravel). SSC decreased dramatically at the end of cascading by exhaustion of easily resuspendable material, as shown by the 500 m and 750 m records (Fig. 2a and Supplementary Fig. 3). SSC was much lower at 200 m depth because of the preferential pathway for dense shelf water and suspended particles along the

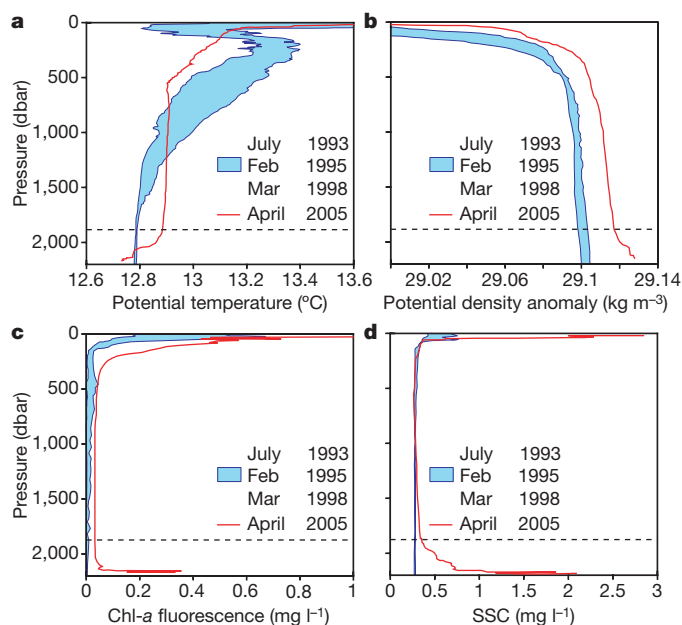


Figure 3 | Deep comparative profiles of key parameters at and off the CCC mouth. **a**, Potential temperature. **b**, Potential density anomaly.

c, Fluorescence. **d**, Suspended sediment concentration. The anomalous April 2005 profiles (red) at the canyon mouth (2,141 m) are compared with the range of normal profiles (blue shaded area) recorded deeper than 2,000 m off the canyon mouth in 1993, 1995 and 1998 (see Methods). The homogeneous water mass observed from 1993 to 1998 corresponds to the Deep Western Mediterranean Water. The April 2005 profiles revealed a large anomaly, with warmer water in the lower half of the water column, resulting from off-shelf dense water formation, overlying a near-bottom, abnormally cold water layer (below dotted line) that demonstrates the intrusion of very dense, chlorophyll-rich and turbid shelf water. See locations in Fig. 1.

southern flank of the canyon (that is, away from the upper canyon axis where the mooring was located), which is in agreement with the development and directions of the giant furrow field. On the basis of these characteristics, the furrow field is interpreted as the seafloor imprint of severe DSWC events repeated through time.

Our observations showing that sediment mass transport affecting large portions of the seafloor can be simply triggered by DSWC add

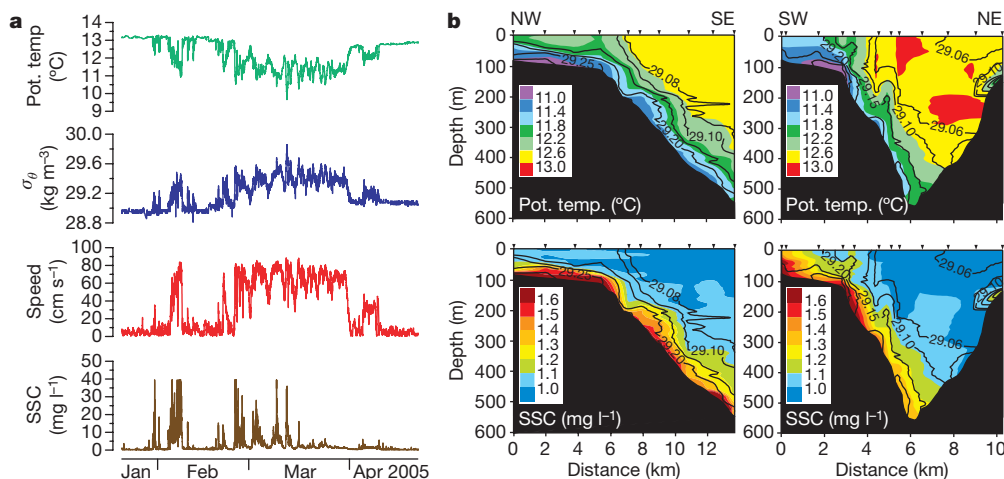


Figure 2 | Time series and sections of key parameters in the CCC.

a, Potential temperature, potential density anomaly, current speed and suspended sediment concentration records at 750 m depth during the cascading period of winter 2004–05. DSWC events correspond to temperature drops concomitant with density increases. **b**, Potential temperature and suspended sediment concentration sections with potential

density anomalies (black contours) along and across the canyon head on 24–26 February 2005. The DSWC plume flows down-canyon along its southern wall. Current meter measurements 5 m above bottom at 750 m depth during the same period recorded down-canyon speeds ranging from 20 to 85 cm s^{-1} . See locations in Fig. 1.

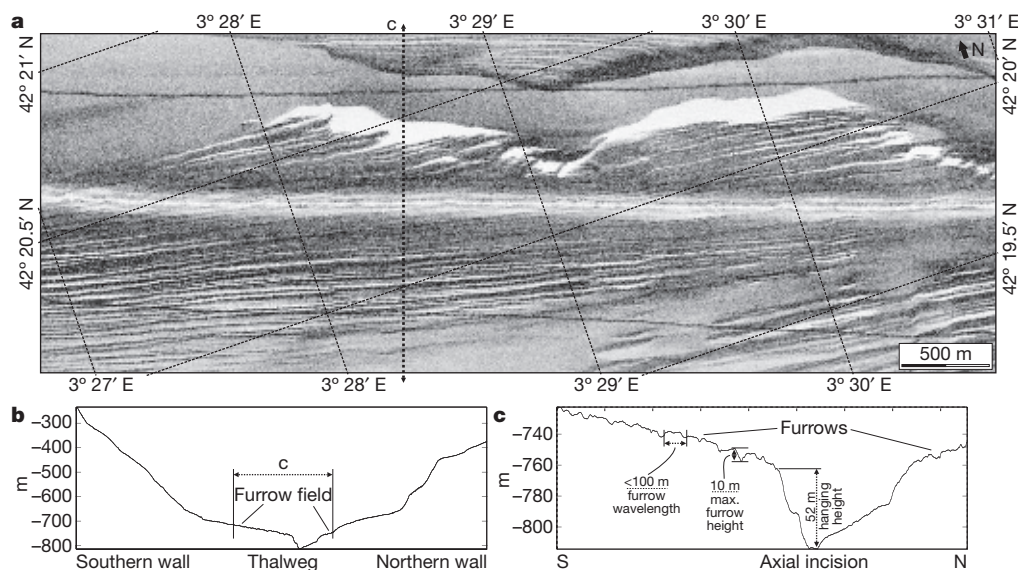


Figure 4 | Acoustic image of the CCC floor. **a**, Side scan sonar high-resolution sonograph of giant furrows on overconsolidated mud. Cohesions up to 40 kPa were measured *in situ* on the same mud type at the inner flank of a nearby canyon. The light grey zone in the lower part of the sonograph suggests that the furrows are locally smoothed or draped by softer

sediment. The lighter band along the middle of the sonograph is an acquisition artefact. **b**, Illustrative section across canyon. **c**, Section across furrow field derived from multibeam bathymetry data showing furrow height (up to 10 m) and wavelength (up to 100 m), and field hanging height (52 m) over the axial incision. See locations in Fig. 1.

an unprecedented dimension to current understanding of sediment gravity flows in the deep sea¹⁵. According to accepted views, river floods and sediment failures initiate most sediment gravity flows including turbidity currents (whose movement is sustained by the upward turbulence of the water–sediment mixture and grain-to-grain interactions), resulting in sediment-laden plumes denser than the surrounding waters because of higher particle concentrations. In the case of DSWC, the water itself is dense enough to sink and rapidly move downwards along the bottom. A $\sim 0.2 \text{ kg m}^{-3}$ horizontal density contrast between the cascade and the ambient water (Fig. 2b) in February 2005 was sufficient to initiate intense downslope water flow without additional sediment load. Whereas computation of theoretical velocities¹⁶ resulted in speeds exceeding 1 m s^{-1} , measurements at 5 m above the bottom revealed speeds as high as 85 cm s^{-1} , equaling those of turbidity currents¹⁷. Its own density and high speed endow DSWC with a strong dragging capacity on loose particles and underconsolidated seafloor sediments.

Beyond sediment transport, DSWC controls off-shelf carbon fluxes and the functioning of deep ecosystems. High concentrations of phytoplankton in cascading waters compared to surrounding waters have been observed elsewhere^{18–20}. The DSWC season in the Gulf of Lions is generally synchronous with high biological production levels in surface waters; the shelf waters showed high chlorophyll-*a* values ($2\text{--}4 \mu\text{g l}^{-1}$) in late February–early March 2005, indicative of a phytoplankton bloom. Water and organic carbon fluxes within the DSWC core have been calculated for the whole cascading period (40 days), using average mean flow speed (0.6 m s^{-1}), plume thickness (60 m) and width (6,000 m). Exported shelf water reached 750 km^3 ($>2/3$ the water volume overlying the Gulf of Lions shelf). With particulate organic carbon (POC) and dissolved organic carbon (DOC) concentrations of 0.1 and 0.7 g m^{-3} in surface waters of the Gulf of Lions, a total organic carbon transport of $0.6 \times 10^6 \text{ t}$ ($15,000 \text{ t d}^{-1}$) can be estimated. Normalizing this cascading transport estimate to the Gulf of Lions' shelf area ($1.2 \times 10^{10} \text{ m}^2$) yields a total organic carbon flux of $50 \text{ g C m}^{-2} \text{ yr}^{-1}$. This is higher than the average export of $16\text{--}25 \text{ g C m}^{-2} \text{ yr}^{-1}$ due to open sea winter convection in the nearby Ligurian Sea²¹. Furthermore, our daily POC export ($1,875 \text{ t d}^{-1}$) is comparable to the upper range of the cascading-driven fluxes estimated for North Atlantic margins^{19–20}.

In promoting such a massive carbon export from the shallowest reservoirs, DSWC contributes to its sequestration in deeper waters where it is less likely to return to the atmosphere. It also directly affects the functioning of the deep ecosystem by providing a fast way of fuelling highly nutritive, fresh organic matter to the deep, as attested by the increased near-bottom water fluorescence recorded at $>2,000 \text{ m}$ depth at the end of the cascading period (Fig. 3c). C/N ratios of sediment trap material caught in canyon heads from November 2003 to April 2004 reflected the shift from degraded organic matter (C/N = 9.5) during autumn stratified conditions to fresh organic matter (C/N = 6.5) during DSWC events (Supplementary Fig. 5).

As many continental margins where DSWC occurs^{8,10} (Supplementary Fig. 6) are cut by deep canyons²², flushing submarine canyons are likely to play a role of global importance in favouring the downslope movement of dense shelf water. Further data collection in DSWC canyoned margins worldwide should be given priority to better constrain global off-shelf sediment and carbon export and the full extent of DSWC impacts on the deep ecosystem.

A final critical point is the fate of DSWC in the coming decades. Modelling results for the Mediterranean Sea anticipate a major decrease of winter deep water formation in the Gulf of Lions²³ following the IPCC-A2 scenario for the twenty-first century as a consequence of a warmer and drier climate over the entire basin²⁴. Deep water formation is also expected to decline in other areas of the world, particularly in high latitudes. In the Atlantic Ocean, the weakening of the thermohaline circulation is anticipated owing to the reduction in dense water formation in Nordic and Arctic regions²⁵. Regionally, such alterations could significantly affect the shelf carbon pump²⁶ and the deep ecosystems whose functioning is linked to DSWC.

METHODS

Our observations result from a unique, dedicated experimental strategy combining adequate space and time (hourly to decadal) multidisciplinary monitoring of on-going processes and high resolution imaging of seafloor features.

Hydrography and organic carbon. Hydrography casts and water sampling were performed using a Seabird 9/11Plus CTD probe, equipped with a 25 cm optical path length Wetlab Cstar transmissometer, a Chelsea ECO/FL fluorometer and a Datasonics altimeter, mounted on a rosette holding Niskin bottles. Water samples were filtered up to filter saturation on pre-weighed $0.45 \mu\text{m}$ Nuclepore

filters. Filters were dried at 40 °C and weighed to derive SSC. POC was analysed in a Leco CN 2000 on the solid residues from water samples filtered on pre-combusted (4 h, 450 °C) glass fibre Whatman GF/F filters (pore size 0.7 µm) and acidified under control with HCl 2 N to remove carbonates. DOC concentration was determined on Whatman GF/F filtered water samples and analysed with a High Temperature Catalytic Oxidation (HTCO) technique on a Shimadzu TOC-V CSH analyser following Cauwet's method²⁷. Background hydrographic information provided in Fig. 3 was derived from the Discovery cruise station MA306 (24 July 1993), the Suivilon 12 cruise station L04 (15 February 1995), and the Fetch cruise station 118 (4 March 1998).

Instrumented moorings. Instrumented moorings deployed at 300 m depth inside canyons during winter 2003–04 (orange dots in Fig. 1) and at 500 m during winter 2004–05 (purple dot in Fig. 1) were equipped with a PPS3 Technicap sequential sediment trap (12 collecting cups) at 30 m above bottom and an Aanderaa RCM9/11 Doppler current meter at 5 m above bottom. 200 and 750 m deep moorings (purple dots in Fig. 1) in the CCC during winter 2004–05 were solely equipped with a near bottom current meter. Current meters were further equipped with temperature, conductivity, pressure and optical backscatter sensors. Sampling intervals were set to 1 week for traps and 20 minutes for current meters. In the LDC, a long term mooring (green dot in Fig. 1) deployed at 1,000 m depth has been in operation since 1993. It is equipped with two PPS3 traps and two Aanderaa RCM7/8 vector-averaging rotor current meters (with temperature and pressure sensors) at 30 m and 500 m above bottom. The long term sampling intervals were set to one month for traps and one hour for current meters.

Sediment samples and sonographs. Bottom sediment samples were collected with a 20 × 30 cm cross-section and 50 cm in height box corer along and across the axis of the CCC, adjacent to the mooring stations at ~200, 500 and 750 m water depth (see location in Fig. 1). Sediment samples were dispersed by a 0.05% sodium metaphosphate solution, wet sieved at 63 µm and analysed on a Sedigraph 5100 for grain-size distribution of the mud fraction. *In situ* undrained shear strength (cohesion) measurements on overconsolidated canyon floor muds were conducted with IFREMER's flexible penetrometer Penfeld. High resolution sonographs along the CCC axis were obtained in 2004 on board R/V *Professor Logachev* using a Mak-1M deep-towed 30 kHz side scan sonar (pulse length 1 ms, horizontal beam width 2°, vertical beam width 60°, TVG 40 dB). This system yielded a total swath range of up to 2 km while towed at a mean height of 100 m over the sea floor at a mean speed of 2.5 knots, with a variable resolution of about 7 to 1 m across and along track. Multibeam bathymetry data from which the section in Fig. 4c was drawn were acquired with an EM-300 Simrad system. Details on the collection and processing of CTD, current meter, sediment trap and side scan sonar data are provided as Supplementary Information.

Received 4 July; accepted 13 September 2006.

- Posamentier, H. W., Jervey, M. T. & Vail, P. R. Eustatic controls on clastic deposition I – Conceptual framework. *Soc. Econ. Paleontol. Miner. Spec. Publ.* **42**, 110–124 (1988).
- Mulder, T., Weber, O., Anschutz, P., Jorissen, F. & Jouanneau, J. M. A few month-old storm generated turbidite deposited in the Capbreton Canyon (Bay of Biscay, SW France). *Geo-Mar. Lett.* **21**, 149–156 (2001).
- Khripounoff, A. *et al.* Direct observation of intense turbidity current activity in the Zaire submarine valley at 4000 m depth. *Mar. Geol.* **194**, 151–158 (2003).
- Paull, C. K. *et al.* Caught in the act: the 20 December 2001 gravity flow event in Monterey Canyon. *Geo-Mar. Lett.* **22**, 227–232 (2003).
- Puig, P. *et al.* Storm-induced sediment gravity flows at the head of the Eel submarine canyon, northern California margin. *J. Geophys. Res.* **109**, C03019, doi:10.1029/2003JC001918 (2004).
- Whitehead, J. A. Dense water off continents. *Nature* **327**, 656 (2002).
- Shapiro, G. I., Huthnance, J. M. & Ivanov, V. V. Dense water cascading off the continental shelf. *J. Geophys. Res.* **108**, 3390, doi:10.1029/2002JC001610 (2003).
- Ivanov, V. V., Shapiro, G. I., Huthnance, J. M., Aleynik, D. L. & Golovin, P. N. Cascades of dense water around the world ocean. *Progr. Oceanogr.* **60**, 47–98 (2004).
- Lacombe, H., Tchernia, P. & Gamberoni, L. Variable bottom water in the Western Mediterranean Basin. *Progr. Oceanogr.* **14**, 319–338, doi:10.1016/00796611(85)90015-1 (1985).
- Durrieu de Madron, X., Zervakis, V., Theocharis, A. & Georgopoulos, D. Comments to "Cascades of dense water around the world ocean". *Progr. Oceanogr.* **64**, 83–90 (2005).
- Béthoux, J. P., Durrieu de Madron, X., Nyffeler, F. & Tailliez, D. Deep water in the western Mediterranean: peculiar 1999 and 2000 characteristics, shelf formation hypothesis, variability since 1970 and geochemical inferences. *J. Mar. Sys.* **33–34**, 117–131 (2002).
- Lopez-Jurado, J. L., Gonzalez-Pola, C. & Velez-Belchi, P. Observation of an abrupt disruption of the long-term warming trend at the Balearic Sea, western Mediterranean Sea, in summer 2005. *Geophys. Res. Lett.* **32**, L24606, doi:10.1029/2005GL024430 (2005).
- Flood, R. D. Classification of sedimentary furrows and a model for furrows initiation and evolution. *Geol. Soc. Am. Bull.* **94**, 630–639 (1983).
- Flood, R. D. Abyssal bedforms as indicators of changing bottom current flow: examples from the U.S. East Coast continental rise. *Paleoceanography* **9**, 1049–1060 (1994).
- Dasgupta, P. Sediment gravity flow—the conceptual problems. *Earth-Sci. Rev.* **62**, 265–281 (2003).
- Shapiro, G. I. & Hill, A. E. Dynamics of dense water cascades at the shelf edge. *J. Phys. Oceanogr.* **27**, 2381–2394 (1997).
- Masson, D. G., Kenyon, N. H. & Weaver, P. P. E. in *Oceanography: An Illustrated Guide* (eds Summerhayes, C. P. & Thorpe, S. A.) 146–151 (Manson, London, 1996).
- Honjo, S., Manganini, S. M. & Wefer, G. Annual particle flux and a winter outburst of sedimentation in the northern Norwegian Sea. *Deep-Sea Res. A* **35**, 1223–1234 (1988).
- Yoder, J. & Ishimaru, T. Phytoplankton advection off the southeastern United States continental shelf. *Cont. Shelf Res.* **9**, 547–553 (1989).
- Hill, A. E. *et al.* The Malin cascade in winter 1996. *J. Mar. Res.* **56**, 87–106 (1998).
- Avril, B. DOC dynamics in the northwestern Mediterranean Sea (DYFAMED site). *Deep-Sea Res. II* **49**, 2163–2182 (2002).
- Earth Resources Information Systems Data Center. Global 30 arc second elevation data set. <<http://edcwww.cr.usgs.gov/landdaac/gtopo30/gtopo30.html>> (1996).
- Somot, S., Sevaut, F. & Deque, M. Transient climate change scenario simulation of the Mediterranean Sea for the twenty-first century using a high-resolution ocean circulation model. *Clim. Dyn.* doi:10.1007/s00382-006-0167-z (published online 20 July 2006).
- Houghton, J. T., *et al.* *Climate Change 2001: The Scientific Basis* (Cambridge Univ. Press, Cambridge, 2001).
- Gregory, J. M. *et al.* A model intercomparison of changes in the Atlantic thermohaline circulation in response to increasing atmospheric CO₂ concentration. *Geophys. Res. Lett.* **32**, L12703, doi:10.1029/2005GL023209 (2005).
- Tsunogai, S., Watanabe, S. & Sato, T. Is there a "continental shelf pump" for the absorption of atmospheric CO₂? *Tellus B* **51**, 701–712, doi:10.1034/j.1600-0889.1999.t01-2-00010.x (1999).
- Cauwet, G. HTCO method for dissolved organic carbon analysis in sea water: influence of catalyst on black estimation. *Mar. Chem.* **47**, 55–64 (1994).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was supported by the European Commission (EUROSTRATAFORM, EURODOM and HERMES projects), the Office of Naval Research, CNRS-INSU and the Catalan Government. Sediment cohesion data were provided by N. Sultan, N data by T. Tesi, and DOC data by M. Pujo-Pay. Multibeam bathymetry was collected in cooperation with Fugro Survey Ltd and AOA Geophysics. Contributions of the scientific and technical staff at the authors' home institutions are warmly acknowledged.

Author Contributions All authors contributed to the design and implementation of the experimental strategy. M.C. steered the integration and joint analysis of the data, interpreted side scan sonar data and wrote the final version of the paper in cooperation with S.H. P.P., X.D.d.M. and A.P. took the responsibility for time series, X.D.d.M. for hydrology data, and S.H. and J.F. for sediment trap data. All authors discussed the results and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to M.C. (miquelcanals@ub.edu).

LETTERS

Predicting the endpoints of earthquake ruptures

Steven G. Wesnousky¹

The active fault traces on which earthquakes occur are generally not continuous¹, and are commonly composed of segments that are separated by discontinuities that appear as steps in map-view. Stress concentrations resulting from slip at such discontinuities may slow or stop rupture propagation and hence play a controlling role in limiting the length of earthquake rupture². Here I examine the mapped surface rupture traces of 22 historical strike-slip earthquakes with rupture lengths ranging between 10 and 420 km. I show that about two-thirds of the endpoints of strike-slip earthquake ruptures are associated with fault steps or the termini of active fault traces, and that there exists a limiting dimension of fault step (3–4 km) above which earthquake ruptures do not propagate and below which rupture propagation ceases only about 40 per cent of the time. The results are of practical importance to seismic hazard analysis where effort is spent attempting to place limits on the probable length of future earthquakes on mapped active faults. Physical insight to the dynamics of the earthquake rupture process is further gained with the observation that the limiting dimension appears to be largely independent of the earthquake rupture length. It follows that the magnitude of stress changes and the volume affected by those stress changes at the driving edge of laterally propagating ruptures are largely similar and invariable during the rupture process regardless of the distance an event has propagated or will propagate.

A concerted effort to construct maps of the geometry of historical earthquake rupture traces commenced around the time of the magnitude-6.4 1968 Borrego Mountain earthquake of California. From these, it became clear that active fault and earthquake rupture traces were often discontinuous and that discontinuities in fault trace were often associated with the endpoints of earthquake ruptures^{1–7}. This in turn led to the idea of a causal association between fault steps and the endpoints of earthquake ruptures and the development of theoretical and numerical models to show the efficacy of the idea^{2,8–14}. Limited examples initially precluded a systematic or statistical analysis of supporting observations. Now there exist more than 20 historical strike-slip surface rupture earthquakes for which maps of the surface trace are available (Table 1). The maps provide a foundation from which to assess more systematically the role of geometrical discontinuities and particularly of fault steps in the propagation of earthquake ruptures.

I focus on continental strike-slip ruptures of length greater than about 15 km. The depth dimension of brittle failure during strike-slip earthquakes is generally limited to about 15 km owing to physical and frictional constraints¹⁵. So the direction of rupture propagation may be viewed as principally horizontal for each event. The approach to data collection and analysis is illustrated in Fig. 1, a map of the 1968 Borrego Mountain surface rupture trace and nearby active fault traces that did not rupture during the earthquake. The location and dimension of fault steps along and at the ends of the earthquake ruptures and the distances to nearest-neighbouring active fault traces from the endpoints of surface rupture traces are annotated. The size of steps in a fault trace are generally taken as the distance between

en echelon strands measured perpendicularly to an average fault strike. Additionally, steps are labelled as restraining or releasing depending on whether volumetric changes within the step resulting from fault slip would yield contractional or dilational strains within the step, respectively².

Thus, for the case of the 1968 Borrego Mountain earthquake, I observe that (1) the rupture propagated across a 1.5 km restraining step, (2) stopped at a 2.5 km restraining step or 7 km releasing step at its northwestern (left) limit, and (3) died at its southeastern (right) limit in the absence of any geometrical discontinuity and at a point where the active trace can be shown to continue uninterrupted for 20 km or more past the end of the rupture. The epicentre is located midway along the north break of the rupture. Maps annotated similarly for the remaining earthquakes in Table 1 are provided in the Supplementary Information.

The resolution of the available maps generally limits observations to discontinuities of about 1 km and greater. Figure 2 serves to illustrate the relationship between the length of rupture and geometrical discontinuities for the earthquakes listed in Table 1. The vertical axis is the distance in kilometres. Along the horizontal axis, I have spaced evenly and ordered by increasing rupture length each of the strike-slip earthquakes listed in Table 1. A dotted line extends vertically from each of the labelled earthquakes. Along each dotted line are plotted various symbols that summarize the size and location of geometrical steps within and at the endpoints of each rupture, as well as where earthquake ruptures have terminated at the ends of active faults. The symbols denote the distance (in kilometres) of steps in surface rupture traces along-strike or the closest distance to the next

Table 1 | Surface rupture earthquakes

Date	Location	Length (km)	M_w
1857 Jan 9	San Andreas, California	360	7.9
1891 Oct 28	Neo-Dani, Japan	80	7.3
1930 Nov 2	Kita-Izu, Japan	35	6.7
1939 Dec 25	Erzincan, Turkey	300	7.7
1940 May 19	Imperial, California	60	6.9
1942 Dec 20	Erbaa Niksar, Turkey	28	6.8
1943 Sep 10	Tottori, Japan	10.5	6.2
1943 Nov 26	Tosya, Turkey	275	7.5
1944 Feb 01	Gerede Bolu, Turkey	135	7.3
1967 Jul 22	Mudurnu, Turkey	60	6.9
1968 Apr 8	Borrego Mountain, California	31	6.1
1979 Oct 15	Imperial, California	36	6.2–6.4
1981 Jul 29	Sirch, Iran	64	6.2
1987 Nov 23	Superstition Hills, California	25	6.2–6.4
1990 Jul 16	Luzon, The Philippines	112	6.9
1992 Jun 28	Landers, California	77	7.2
1998 Mar 14	Fandoqa, Iran	25	6.6
1999 Aug 17	Izmit, Turkey	145	7.1
1999 Oct 16	Hector Mine, California	44	6.9
1999 Nov 12	Duzce, Turkey	40	7.0
2001 Nov 14	Kunlun, China	421	7.8
2002 Nov 03	Denali, Alaska	302	7.6

Sources used to construct fault trace maps used in development of Figs 2 and 3 of the manuscript are provided in the Supplementary Information.

¹Center for Neotectonic Studies, Mail Stop 169, University of Nevada, Reno, Nevada 89557, USA.

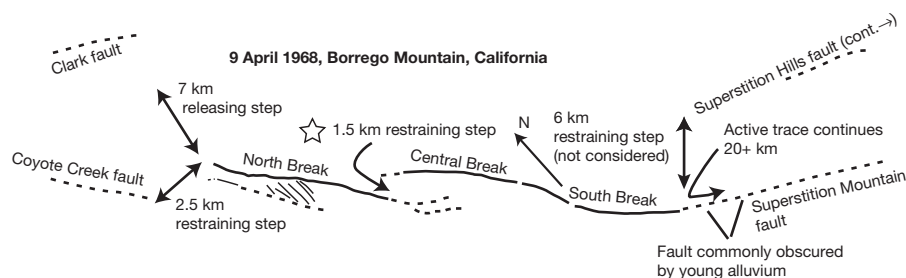


Figure 1 | Map of 1968 Borrego Mountain earthquake surface trace. Adjacent and continuing traces of active faults that did not rupture during the earthquake are shown as dotted lines. Also annotated are the dimensions

mapped active fault from the terminus of the respective ruptures. Separate symbols are used according to whether the steps are releasing or restraining in nature, and whether they occur within (green open symbols: rupture continues through) or at the endpoints of the rupture trace (red solid symbols). In certain instances, the endpoints of rupture are not associated with a discontinuity in fault strike, in which case the endpoint of rupture is denoted by a separate symbol (open orange circles) and annotated with the distance that the active trace continues beyond the endpoint of rupture. Because of the complexity of some ruptures and presence of subparallel and branching fault traces, some earthquakes have more than two 'ends'.

The map of the 1968 earthquake rupture shown in Fig. 1 serves as an example of the approach and the data presentation in Fig. 2, where it is synoptically shown that the fault ruptured through a 1.5 km restraining step, stopped on one end at either a 2.5 km restraining step or 7 km releasing step, and stopped at the other end along an active trace that continues for 20 km or more in the absence of any

of fault steps measured approximately perpendicular to fault strike and the distance to the nearest-neighbouring fault from the 1968 rupture endpoints. The star is the earthquake epicentre.

observable discontinuity. The colour scheme of symbols follows that of a conventional red-yellow-green stop light: Ruptures appear to have ended at the discontinuities coloured red, jumped across the discontinuities coloured green, and simply died out along-strike in the absence of any discontinuities for the cases coloured yellow.

The compilation of observations summarized in Fig. 2 indicates that about two-thirds of terminations of strike-slip ruptures are associated with geometrical steps in fault trace or the termination of the active fault on which they occurred. The histogram in Fig. 3 gives a statistical idea of whether or not an earthquake rupture will be associated with a particular fault step or fault terminus. The observations of discontinuities are binned as a function of their size and coded according to whether they sit within or at the terminus of an earthquake rupture. The plot shows that a transition exists between 3 and 4 km, above which rupture fronts have not been observed to propagate through. At steps of lesser dimension ruptures appear to cease propagating only about 40% of the time. The result may be of

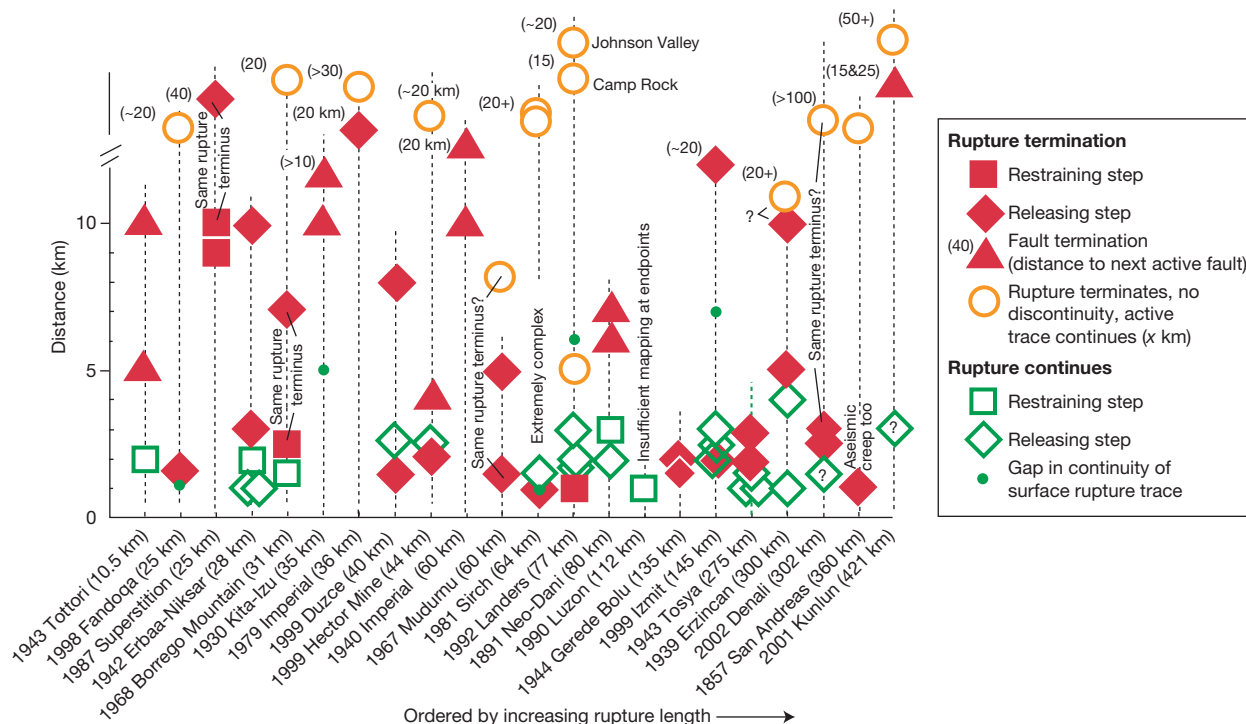


Figure 2 | Synopsis of observations bearing on relationship of geometrical discontinuities along fault strike to the endpoints of historical earthquake ruptures. Earthquake date, name and rupture length listed on horizontal axis. The earthquakes are ordered by increasing rupture length (but not scaled to distance along-axis). Above the label of each earthquake is a vertical line and symbols along the line represent dimension of discontinuities within and at endpoints of each rupture. The dimension of discontinuities are measured as distance across fault step approximately perpendicular to

fault strike or the distance from rupture terminus to nearest-neighbouring active fault trace. Discontinuities through which ruptures passed (broke through) are green open squares and diamonds. Discontinuities located at ends of ruptures are red solid squares, triangles and diamonds. The orange open circles represent the endpoints or earthquake ruptures which are not associated with a geometrical discontinuity and the value next to each is an indicator of the distance along which the active trace continues past the endpoint of the historical rupture.

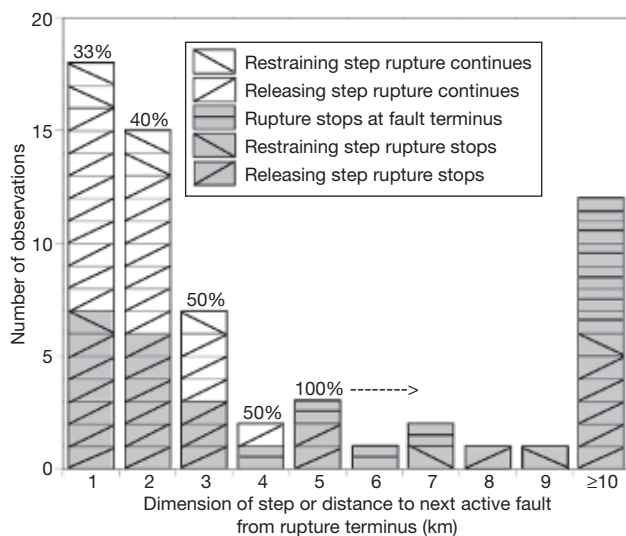


Figure 3 | Geometrical discontinuities as a function of size. Histogram of the total number of geometrical discontinuities located along historical strike-slip ruptures binned as a function of size (≥ 1 , ≥ 2 , ...) and shaded according to whether the particular step occurred at the endpoint of rupture (shaded) or was broken through by the rupture (not shaded).

practical importance in seismic hazard analysis where effort is spent attempting to place limits on the probable length of future earthquakes on mapped active faults. I surmise that the variability of behaviour for steps of dimension less than 3–4 km in part reflects variability in the three-dimensional character of the discontinuities mapped at the surface. The effect on rupture propagation may vary between steps of equal map dimension if, for example, the subsurface structures differ or do not extend to equal depths through the seismogenic layer^{16,17}.

Turning back to Fig. 2, I further note that the transition mentioned above seems largely independent of rupture length. I therefore suggest that the magnitude of stress changes and the volume effected by those stress changes at the leading edge of propagating earthquake ruptures are similar at the initial stages of rupture propagation and largely invariable during the rupture process. In this context, it appears that variations in earthquake rupture lengths are not necessarily controlled by the relative size of initial slip pulses or stress drops^{18,19} but rather by the geometrical complexity of fault traces¹ and variations in accumulated stress levels along faults that arise owing to the location of past earthquakes along the respective faults²⁰.

Finally, I mention recent theoretical work that implies that releasing steps should be easier to rupture through than restraining steps^{9,11,12} although it is releasing steps that are observed more frequently at the endpoints of ruptures (Fig. 2). The apparent conflict resides in the observation that releasing steps outnumber restraining steps by six to one in the data set. Restraining steps may be relatively more efficient at impeding individual earthquake ruptures, but the observations indicate that deformation processes attendant on cumulative strike-slip displacements are relatively more efficient in

the creation and maintenance of extensional steps and analogously in the linkage and removal of restraining steps.

Received 23 May; accepted 21 September 2006.

- Wesnousky, S. Seismological and structural evolution of strike-slip faults. *Nature* **335**, 340–343 (1988).
- Segall, P. & Pollard, D. D. Mechanics of discontinuous faults. *J. Geophys. Res.* **85**, 4337–4350 (1980).
- King, G. C. P. & Nabelek, J. F. Role of fault bends in the initiation and termination of earthquake rupture. *Science* **228**, 984–987 (1985).
- Barka, A. & Kadinsky-Cade, K. Strike-slip fault geometry in Turkey and its influence on earthquake activity. *Tectonics* **7**, 663–684 (1988).
- Schwartz, D. P. & Sibson, R. H. (eds) *Fault Segmentation and Controls of Rupture Initiation and Termination* (United States Geological Survey, USGS open-file report 89-315, Proc. Conf. XLV, Palm Springs, California, 1989).
- Zhang, P., Slemmons, D. B. & Mao, F. Geometric pattern, rupture termination, and fault segmentation of the Dixie Valley-Pleasant Valley active normal fault systems, Nevada. *USA J. Struct. Geol.* **13**, 165–176 (1991).
- Sibson, R. H. Stopping of earthquake ruptures at dilational fault jogs. *Nature* **316**, 248–251 (1985).
- Sibson, R. H. in *Earthquake Source Mechanics* 157–167 (American Geophysical Union, Washington DC, 1986).
- Harris, R. A. & Day, S. M. Dynamics of fault interaction—parallel strike-slip faults. *J. Geophys. Res.* **18**, 4461–4472 (1993).
- Harris, R. A. & Day, S. M. Dynamic 3D simulations of earthquakes on en echelon faults. *Geophys. Res. Lett.* **98**, 2089–2092 (1999).
- Oglesby, D. D. The dynamics of strike-slip step-overs with linking dip-slip faults. *Bull. Seismol. Soc. Am.* **95**, 1604–1622 (2005).
- Harris, R. A., Archuleta, R. J. & Day, S. M. Fault steps and the dynamic rupture process: 2-D numerical simulations of a spontaneously propagating shear fracture. *Geophys. Res. Lett.* **18**, 893–896 (1991).
- Harris, R. A., Dolan, J. F., Hartleb, R. & Day, S. M. The 1999 Izmit, Turkey earthquake—A 3D dynamic stress transfer model of intraequake triggering. *Bull. Seismol. Soc. Am.* **92**, 245–255 (2002).
- Kase, Y. & Kuge, K. Numerical simulation of spontaneous rupture processes on two non-coplanar faults: the effect of geometry on fault interaction. *Geophys. J. Int.* **135**, 911–922 (1998).
- Brace, W. F. & Kohlstedt, D. L. Limits on lithospheric stress imposed by laboratory experiments. *J. Geophys. Res.* **85**, 6248–6252 (1980).
- Graymer, R. W., Langenheim, V. E., Simpson, R. W., Jachens, R. C. & Ponce, D. A. Relatively simple throughgoing fault planes at large earthquake depth may be concealed by surface complexity in stepover regions. In *Tectonics of Strike-slip Restraining and Releasing Bends* (eds Cunningham, D. & Mann, P.) (Geological Society of London Special Volume, in the press).
- Simpson, R. W., Barall, M., Langbein, J., Murray, J. R. & Rymer, M. J. San Andreas fault geometry in the Parkfield, California region. *Bull. Seismol. Soc. Am.* **96**, S28–S37 (2006).
- Bodin, P. & Brune, J. N. On the scaling of slip with rupture length for shallow strike-slip earthquakes: Quasi-static models and dynamic rupture propagation. *Bull. Seismol. Soc. Am.* **86**, 1292–1299 (1996).
- Heaton, T. H. Evidence for and implication of self-healing pulses of slip in earthquake rupture. *Phys. Earth Planet. Inter.* **64**, 1–20 (1990).
- McCann, W. R., Nishenko, S. P., Sykes, L. R. & Krause, J. Seismic gaps and plate tectonics: seismic potential for major boundaries. *Pure Appl. Geophys.* **117**, 1082–1147 (1979).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements I thank J. Dolan, R. Dmowska, R. Harris, B. Oglesby and B. Shaw for comments or reviews when developing the manuscript. Research was supported in part by a USGS NHERP contract and an NSF/SCEC award.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The author declares no competing financial interests. Correspondence and requests for materials should be addressed to the author (stevew@seismo.unr.edu).

Allee effects and pulsed invasion by the gypsy moth

Derek M. Johnson¹, Andrew M. Liebhold², Patrick C. Tobin² & Ottar N. Bjørnstad³

Biological invasions pose considerable threats to the world's ecosystems¹ and cause substantial economic losses². A prime example is the invasion of the gypsy moth in the United States, for which more than \$194 million was spent on management and monitoring between 1985 and 2004 alone³. The spread of the gypsy moth across eastern North America is, perhaps, the most thoroughly studied biological invasion in the world, providing a unique opportunity to explore spatiotemporal variability in rates of spread. Here we describe evidence for periodic pulsed invasions, defined as regularly punctuated range expansions interspersed among periods of range stasis. We use a theoretical model with parameter values estimated from long-term monitoring data to show how an interaction between strong Allee effects (negative population growth at low densities)⁴ and stratified diffusion (most individuals disperse locally, but a few seed new colonies by long-range movement)⁵ can explain the invasion pulses. Our results indicate that suppressing population peaks along range borders might greatly slow invasion.

Since its accidental release near Boston, Massachusetts in 1869, the Eurasian gypsy moth (*Lymantria dispar*) has invaded more than 1,000,000 km² of the northeastern United States⁵. Despite an extensive containment programme, gypsy moth invasion continues, causing large-scale defoliation (700–50,000 km² annually) and, occasionally, extensive tree mortality in eastern deciduous forests⁶. Currently, the gypsy moth's North American range extends from Maine to North Carolina and west into Wisconsin^{7,8} (Fig. 1). Gypsy moth outbreaks are cyclic in the United States, as they are in its native Eurasian range, and the population undergoes large fluctuations, usually with a 10-yr period^{9,10}. The historical rate of gypsy moth invasion in the northeastern

United States averages around 21 km yr⁻¹ (ref 5). The temporal variations in invasion rates are spatially synchronized across distances of up to 600 km (Fig. 2). Since 1988, the US Department of Agriculture has implemented a containment programme to suppress nascent populations at the frontier of the gypsy moth's expanding range⁷. The programme seems to have slowed the spread, but the gypsy moth still threatens to establish itself throughout most of the United States¹¹.

Diffusion models provide a natural starting point for understanding invasions. Skellam's classic model¹² showed that the radial rate of range expansion (V) is a joint function of the intrinsic rate of population increase (r) and the dispersal rate (D), according to $V = 2(rD)^{0.5}$. The key insight of this model is that the velocity of invasion should be constant when r and D are constant. Recent theories have predicted that the presence of multiple dispersal modes can result in accelerated invasions¹³, and that environmental fluctuations can cause variations in rates of spread¹⁴.

Adult female gypsy moths are flightless, and ballooning of 1st instars usually occurs only over short distances. Long-range movement is most often due to anthropogenic transportation of life stages, providing a second mode of spread. This leads to stratified diffusion¹⁵. Our analyses of historical county-level quarantine data between 1960 and 2002 show periodically fluctuating rates of invasion at pulses with a significant 4-yr period ($P < 0.05$; Fig. 3). We propose that these invasion pulses are caused by an interaction between stratified diffusion and a strong Allee effect.

Strong Allee effects can influence species ranges by prohibiting range expansion, termed 'invasion pinning'¹⁶. In invasion pinning, dispersal into habitats beyond the range border is too low to inflate population densities above the Allee threshold; therefore, nascent populations beyond the established range fail and, consequently, so does range expansion. However, when the number of immigrants exceeds the Allee threshold—the minimum density required for a

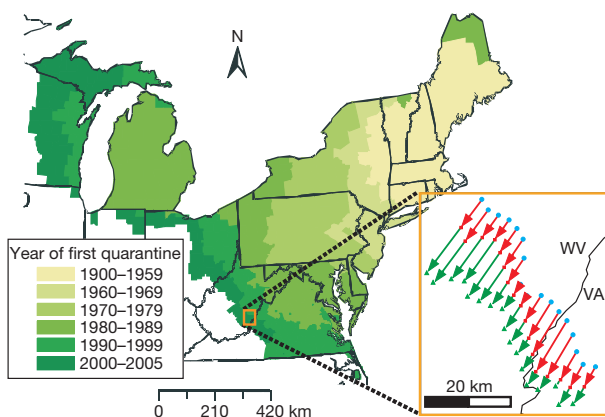


Figure 1 | Map showing the years of first quarantine of the gypsy moth in the eastern United States. The insert shows gypsy moth spread as measured by population isocline displacement from year to year (years represented by different colours) when measured along perpendicular transects spaced ~6 km apart. Each arrow therefore represents the estimated range expansion from one year to the next at a specific location. VA, Virginia; WV, West Virginia.

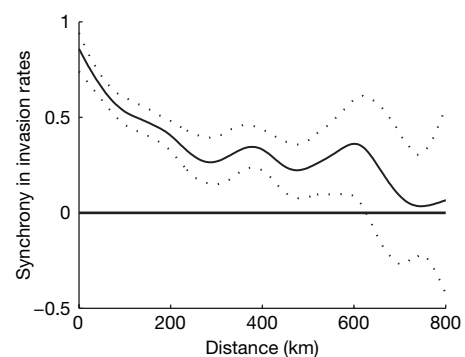


Figure 2 | Nonparametric spatial correlation function showing geographically synchronized rates of invasion out to a distance of 600 km. The full line represents the estimated distance-dependent synchrony and the dotted lines represent the 95% bootstrap confidence limits.

¹Department of Biology, University of Louisiana, PO Box 42451, Lafayette, Louisiana 70504, USA. ²United States Department of Agriculture, Forest Service, Northern Research Station, 180 Canfield Street, Morgantown, West Virginia 26505, USA. ³501 ASI Building, Departments of Entomology and Biology, Pennsylvania State University, University Park, Pennsylvania 16802, USA.

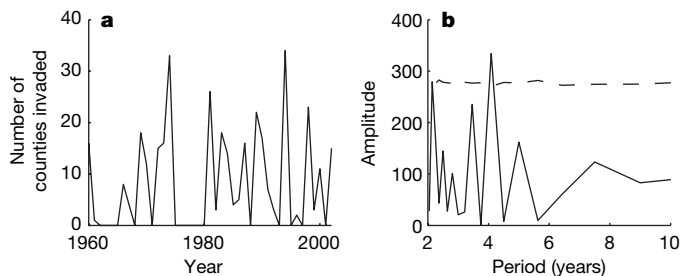


Figure 3 | Periodic invasion pulses in the gypsy moth. **a**, The number of counties quarantined for gypsy moth in each year from 1960 to 2002. **b**, The periodogram (solid line) reveals a significant 4-yr period in the rate of invasion by the gypsy moth. The dashed line represents the 95% null distribution from the randomization test.

population to grow—budding populations attain a positive growth rate and can establish.

Allee effects have been recognized as important in theoretical^{16–18} and empirical^{19,20} studies of invasions²¹. There are several studies that show Allee effects at low densities of gypsy moths^{22–24}, but estimates of the Allee threshold are usually approximate because most counts are zero, and high observation error and demographic stochasticities are inevitable consequences of low abundance. The current gypsy moth containment programme offers an exception because of its extensive grids of pheromone-baited traps (>100,000 traps per year), which are sensitive to extremely low moth densities along the invasion front. The result is what we believe to be the clearest illustration, and the most detailed quantification, of an Allee threshold in the gypsy moth (~17 moths per trap). The programme also allows us to estimate a ‘carrying capacity’. The notion of a carrying capacity is nebulous for widely fluctuating species like the gypsy moth because it is unlikely to represent a stable attractor. However, with a working definition of carrying capacity as the upper threshold, below which populations are more likely to increase than decrease and above which the reverse is true, we pinpoint the ‘carrying capacity’ at around 687 moths per trap (the converse is our working definition of the Allee threshold; see Supplementary Information A), and were able to estimate the parameters of our model.

With stratified diffusion¹⁵, large-scale range expansion is initiated by the seeding of nascent colonies that are isolated from the current range. These are usually of low abundance, particularly when the donor population is sparse. Without a strong Allee effect, these nascent populations will grow and, according to standard theory, invaders should establish—possibly with an allowance for resource fluctuations²⁵—from any number of colonizers. When models lack an Allee effect or stratified diffusion, the simulated invasion rate also lacks periodicity (see Supplementary Information B). By contrast, strong Allee effects allow nascent populations to grow and invasion to progress only if the number of colonizers exceeds the Allee threshold. Thus, range expansion can resume only when the donor populations have grown large enough—the ‘donor threshold’—to provide sufficient numbers of emigrants for further spread. Strong Allee effects, therefore, might induce donor thresholds and result in punctuated and pulsed invasions. The frequency of pulsing is determined by the time required for the population density in a newly colonized location to exceed the donor threshold.

These conclusions stem from a spatially explicit model on a one-dimensional discrete landscape of local populations that grow according to a 2nd-order Moran–Ricker model^{26,27} (parameterized to reflect key features of gypsy moth dynamics), and spread according to an exponential dispersal kernel but with occasional long-range dispersal (‘stratified diffusion’). With an Allee threshold and post-invasion fluctuations around the carrying capacity mirroring that observed in the data, and with stratified diffusion, invasion is predicted to be pulsed with a 4-yr period (Fig. 4). This matches the periodicity observed in the 1960–2002 quarantine records. We

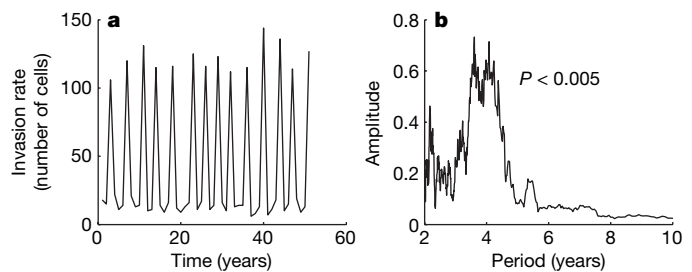


Figure 4 | Simulation model. **a**, Time series of invasion rates (number of cells invaded) during a theoretical invasion in the model with an Allee effect and stratified diffusion. **b**, The power of periodicity, calculated from spectral analysis, of invasion pulses show a peak around 4 yr. The periodicity was smoothed by averaging moving windows of 10 periodicities.

increased dispersal distance to more closely approximate continuous space, and showed that periodic pulses are not artefacts of the discrete-space model (see Supplementary Information C). Moreover, a simplified, continuous-space (point-process) model confirms periodic pulses of invasion in a model with an Allee effect (see Supplementary Information D). Sensitivity analyses show that the prediction of periodic invasion pulses is robust against moderate variation in the model parameters, and indicates that first- or second-order density dependence, the magnitude of environmental stochasticity, or the presence/absence of population cycles has no qualitative effects on the prediction (see Supplementary Information E). These results indicate that stratified diffusion might be a key ingredient in the production of invasion pulses, a conclusion that is consistent with the previously published models with Allee effects, but without stratified diffusion, that lack invasion pulses¹⁸.

The time for a nascent population to increase to the donor threshold will depend on three factors: the magnitude of the Allee threshold, the growth rate of a population at low densities, and the emigration rate. The invasion period observed here (4–5 yr) is approximately half of the outbreak cycle (9–10 yr). Thus, the time for newly established populations to grow sufficiently large to serve as donor populations is similar to the time for a resident population to grow from low to outbreak levels, leading us to hypothesize that the donor threshold is crossed as populations near the expanding population front approach outbreak levels. In the model, invasion pulses are not synchronized with regional gypsy moth outbreaks because newly colonized populations are initially asynchronous with long-established populations. But invasion pulses are themselves spatially synchronous (Fig. 2), indicating that the growth of these populations is synchronized along the invasion front. The extensive data on gypsy moth populations might provide an excellent opportunity to test this prediction.

Understanding the processes that underlie observed patterns is key to devising successful management plans for invasive species. Slowing the spread of the gypsy moth is a priority in forest management in the United States. The current containment programme^{7,8} aims to eradicate new populations beyond the established invasion front. Our results indicate that the invasion might also be slowed by suppressing outbreaks near the invasion front, to reduce the number of dispersers to below the donor threshold. The robustness of our model’s prediction of periodic invasion pulses in the face of variation in most parameters except for the Allee effect is intriguing; if Allee effects and stratified diffusion are important in other invaders, it might be of interest to look for their footprint in patterns of pulsed invasion. Wherever they are found, suppression of donor population densities could be as important as extinguishing nascent foci to the control of a biological invasion.

METHODS

The data. The data are, in part, from the historical, county-level, gypsy moth quarantine status (US Code of Federal Regulations, Title 7, Chapter III, Section 301.45-3) as reported by the US Department of Agriculture (USDA) since 1934 and compiled in a geographical information system⁵. The spread has been

through expansion into contiguous counties with the exception of a disjunct population in central Michigan (excluded in our analysis).

As a part of the effort to slow the gypsy moth's spread across the United States, intensive monitoring efforts have been focused around the invasion front. Every year, around 150,000 pheromone-baited traps are deployed across an area that currently covers 344,000 km² (ref. 8). In central Virginia and West Virginia, data were collected annually from 1988 to 2004. Each year, more than 13,000 traps were placed at 2-, 3- or 8-km inter-trap distances up to 150 km from the border of the gypsy moth invasion front⁸. The trap data were transformed using $\log_{10} + 1$, and interpolated at a 1-km scale using median indicator kriging²⁸. The resulting smoothed surface for each year was used to estimate isoclines of 1, 3, 10, 30 and 100 moths per trap. Local invasion rates were subsequently calculated by measuring the isocline displacements in consecutive years at fixed intervals²⁹ (Fig. 1) along the isoclines (see Supplementary Information A).

The Allee threshold and 'carrying capacity' were estimated from the pheromone-baited trap data by quantifying the patterns of increase and decrease in trap capture for each year (see Supplementary Information B).

The model. The model is a spatially extended, stochastic, second-order Moran–Ricker model^{26,27}—represented here on the natural scale as opposed to the common log-linear formulation²⁶—with an Allee effect:

$$n_{i,t} = \sum_{j=-50}^{i+50} \left[(a n_{i,t-1}) (e_i^v) (n_{i,t-1}^\alpha n_{i,t-2}^\beta) \left(\frac{n_{i,t-1}^2}{c^2 + n_{i,t-2}^2} \right) (d_{ij}^\tau) \right] + n_{i-k,t-1} \phi \quad (1)$$

Here, $n_{i,t}$ is the abundance at location i in generation t and a is the maximum population growth rate. We pinpoint this parameter at 26.3 because, along with the specific parameters of density-dependence (see below), this makes the model mirror the post-invasion fluctuations around 687 captures per trap that are seen in the historical data (see Supplementary Information A). The second term, e_i^v , represents unit-mean, log-normally distributed environmental stochasticity in population growth. On the basis of the general understanding that such stochasticity arises from a combination of local (σ) and regional (ρ) processes, we model v as the mixture $\rho \rho_{i,t} + (1 - \rho) \sigma_t$, where $\rho_{i,t}$ and σ_t represent zero-mean random Gaussian variates each with a variance u_i and r represents the relative importance of the local versus regional variability. In our analyses we set the regional versus local stochasticity at 75%:25%²⁷ and arbitrarily set u at 0.4. The parameters α and β represent the strength of first- and second-order density-dependence, respectively. These have previously been estimated at $\alpha = -0.1$ and $\beta = -0.4$ from gypsy moth time-series data²⁷, reflecting the noisy 10-yr cycle of gypsy moth outbreaks. The fourth term represents a sigmoid Allee effect, for which $c = 0$ denotes no Allee effect. We use the empirical value of $c = 39.4$, which produces an Allee threshold of 17 individuals in the model (see Supplementary Information A). In the simulation, we considered a one-dimensional landscape consisting of 1,000 locations. Each location cell was linked through dispersal according to an arbitrarily scaled exponential kernel ($\tau = -0.19$) with a maximum local dispersal distance of 100 cells (d_{ij} is the distance between cells i and j). Jump dispersal, according to a stratified diffusion process^{13,15}, has a proportion of the population (ϕ)—arbitrarily set at 0.005—jump a distance (k) as randomly selected from a uniform distribution on the integers [100, ..., 200]. Sensitivity analyses of all parameters are provided in Supplementary Information E.

Spatial analysis. Isocline displacements in all pairs of consecutive years, from 1988 to 2004, were used to quantify the spatial synchrony of invasion rates at stationary spatial locations (Fig. 1). The spatial correlation of invasion rates in the time series was analysed using a nonparametric spatial correlation function³⁰ in R. Confidence intervals were constructed using bootstrap resampling based on 500 iterations.

Spectral analysis. Periodograms were used to identify periodicity in both the empirical data and simulated models of invasion dynamics. With respect to the Allee effect, we defined an invasion pulse as when a nascent population beyond the invasion front first attains a positive growth rate (excluding migration). For the empirical data, significance levels for powers of periodicity were calculated through randomization tests using 10,000 permutations. Because the spectral analyses on the model (Fig. 4, Supplementary Information B) were calculated from a simulation with ~5,000 invasion pulses, the spectral powers were smoothed with moving averages across windows of 10 periods (the windows covered a very small periodicity range, especially at periodicities below 10 time steps). Significant periodicities were assessed by comparing the maximum power of periodicity to a distribution of maximum powers of periodicity from 10,000 resamplings of the time series with replacement.

Received 23 May; accepted 11 September 2006.

1. Vitousek, P. M., D'Antonio, C. M., Loope, L. L. & Westbrooks, R. Biological invasions as global environmental change. *Am. Sci.* **84**, 468–478 (1996).
2. Barbier, E. B. A note on the economics of biological invasions. *Ecol. Econ.* **39**, 197–202 (2001).

3. Mayo, J. H., Straka, T. J. & Leonard, D. S. The cost of slowing the spread of the gypsy moth (Lepidoptera: Lymantriidae). *J. Econ. Entomol.* **96**, 1448–1454 (2003).
4. Allee, W. C. *The Social Life of Animals* (Norton, New York, 1938).
5. Liebhold, A. M., Halverson, J. A. & Elmes, G. A. Gypsy moth invasion in North America: a quantitative analysis. *J. Biogeogr.* **19**, 513–520 (1992).
6. Montgomery, M. E. & Wallner, W. E. in *Dynamics of Forest Insect Populations: Patterns, Causes, Implications* (ed. Berryman, A. A.) 253–375 (Plenum, New York, 1988).
7. Sharov, A. A., Leonard, D., Liebhold, A. M., Roberts, E. A. & Dickerson, W. "Slow the Spread": A national program to contain the gypsy moth. *J. For.* **100**, 30–35 (2002).
8. Tobin, P. C. *et al.* Management of the gypsy moth through a decision algorithm under the Slow-the-Spread project. *Am. Entomol.* **50**, 200–209 (2004).
9. Dwyer, G., Dushoff, J. & Yee, S. H. The combined effects of pathogens and predators on insect outbreaks. *Nature* **430**, 341–345 (2004).
10. Johnson, D. M., Liebhold, A. M. & Bjørnstad, O. N. Circumpolar variation in periodicity and synchrony among gypsy moth populations. *J. Anim. Ecol.* **74**, 882–892 (2005).
11. Sharov, A. A. & Liebhold, A. M. Bioeconomics of managing the spread of exotic pest species with barrier zones. *Ecol. Appl.* **8**, 833–845 (1998).
12. Skellam, J. G. Random dispersal in theoretical populations. *Biometrika* **38**, 196–218 (1951).
13. Shigesada, N., Kawasaki, K. & Takeda, Y. Modeling stratified diffusion in biological invasions. *Am. Nat.* **146**, 229–251 (1995).
14. Neubert, M. G., Kot, M. & Lewis, M. A. Invasion speeds in fluctuating environments. *Proc. R. Soc. Lond. B* **267**, 1603–1610 (2000).
15. Sharov, A. A. & Liebhold, A. M. Model of slowing the spread of gypsy moth (Lepidoptera: Lymantriidae) with a barrier zone. *Ecol. Appl.* **8**, 1170–1179 (1998).
16. Keitt, T. H., Lewis, M. A. & Holt, R. D. Allee effects, invasion pinning, and species' borders. *Am. Nat.* **157**, 203–216 (2001).
17. Wang, M. H. & Kot, M. Speeds of invasion in a model with strong or weak Allee effects. *Math. Biosci.* **171**, 83–97 (2001).
18. Lewis, M. A. & Kareiva, P. Allee dynamics and the spread of invading organisms. *Theor. Popul. Biol.* **43**, 141–158 (1993).
19. Taylor, C. M., Davis, H. G., Cville, J. C., Grevstad, F. S. & Hastings, A. Consequences of an Allee effect in the invasion of a Pacific estuary by *Spartina alterniflora*. *Ecology* **85**, 3254–3266 (2004).
20. Veit, R. R. & Lewis, M. A. Dispersal, population growth, and the Allee effect: Dynamics of the House Finch invasion of eastern North America. *Am. Nat.* **148**, 255–274 (1996).
21. Taylor, C. M. & Hastings, A. Allee effects in biological invasions. *Ecol. Lett.* **8**, 895–908 (2005).
22. Liebhold, A. & Bascompte, J. The Allee effect, stochastic dynamics and the eradication of alien species. *Ecol. Lett.* **6**, 133–140 (2003).
23. Sharov, A. A., Liebhold, A. M. & Ravlin, F. W. Prediction of gypsy-moth (Lepidoptera, Lymantriidae) mating success from pheromone trap counts. *Environ. Entomol.* **24**, 1239–1244 (1995).
24. Whitmore, S. L. & Tobin, P. C. Persistence of invading gypsy moth colonies in the United States. *Oecologia* **147**, 230–237 (2006).
25. Dwyer, G. & Morris, W. F. Resource-dependent dispersal and the speed of biological invasions. *Am. Nat.* **167**, 165–176 (2006).
26. Royama, T. *Analytical Population Dynamics* (Chapman & Hall, London, 1992).
27. Peltonen, M., Liebhold, A. M., Bjørnstad, O. N. & Williams, D. W. Spatial synchrony in forest insect outbreaks: roles of regional stochasticity and dispersal. *Ecology* **83**, 3120–3129 (2002).
28. Deutsch, C. V. & Journel, A. G. *GSLIB: Geostatistical Software Library and User's Guide* (Oxford Univ. Press, New York, 1992).
29. Sharov, A. A., Roberts, E. A., Liebhold, A. M. & Ravlin, R. W. Gypsy moth (Lepidoptera: Lymantriidae) spread in the central Appalachians: three methods for species boundary estimation. *Environ. Entomol.* **24**, 1529–1538 (1995).
30. Bjørnstad, O. N. & Falck, W. Nonparametric spatial covariance functions: estimation and testing. *Environ. Ecol. Stat.* **8**, 53–70 (2001).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank L. Blackburn for invaluable assistance in manuscript preparation. B. Grenfell provided insightful comments that improved the manuscript. This work was supported by the National Research Initiative of the USDA Cooperative State Research, Education and Extension Service Grants to O.N.B. and A.M.L. (2002), O.N.B., A.M.L. and P.C.T. (2006), and D.M.J. (2006).

Author Contributions The original concept of the manuscript was conceived by A.M.L., O.N.B. and D.M.J., and all authors contributed to the development of that concept. A.M.L. provided expertise regarding the ecology of the gypsy moth. P.C.T. performed the analyses for estimates of the Allee effect and carrying capacity. A.M.L. and O.N.B. constructed the population equation, D.M.J. wrote the code for the spatial model and ran the spatial simulations, and performed tests for periodicity and sensitivity analysis. O.N.B. coded and ran the continuous-space model. D.M.J. was responsible for writing the manuscript, and all authors contributed equally to revisions.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to D.M.J. (derekjohnson@louisiana.edu).

LETTERS

A protein interaction network for pluripotency of embryonic stem cells

Jianlong Wang¹, Sridhar Rao¹, Jianlin Chu¹, Xiaohua Shen¹, Dana N. Levasseur¹, Thorold W. Theunissen¹ & Stuart H. Orkin^{1,2}

Embryonic stem (ES) cells are pluripotent^{1,2} and of therapeutic potential in regenerative medicine^{3,4}. Understanding pluripotency at the molecular level should illuminate fundamental properties of stem cells and the process of cellular reprogramming. Through cell fusion the embryonic cell phenotype can be imposed on somatic cells, a process promoted by the homeodomain protein Nanog⁵, which is central to the maintenance of ES cell pluripotency^{6,7}. Nanog is thought to function in concert with other factors such as Oct4 (ref. 8) and Sox2 (ref. 9) to establish ES cell identity. Here we explore the protein network in which Nanog operates in mouse ES cells. Using affinity purification of Nanog under native conditions followed by mass spectrometry, we have identified physically associated proteins. In an iterative fashion we also identified partners of several Nanog-associated proteins (including Oct4), validated the functional relevance of selected newly identified components and constructed a protein interaction network. The network is highly enriched for nuclear factors that are individually critical for maintenance of the ES cell state and co-regulated on differentiation. The network is linked to multiple co-repressor pathways and is composed of numerous proteins whose encoding genes are putative direct transcriptional targets of its members. This tight protein network seems to function as a cellular module dedicated to pluripotency.

To recover Nanog with its associated proteins, we used a biotinylation/proteomics approach¹⁰. Nanog complementary DNA bearing an amino-terminal Flag epitope and a short peptide tag that serves as a substrate for *in vivo* biotinylation (Fig. 1a) was expressed at about 20% of the endogenous level in ES cells previously engineered to express *Escherichia coli* biotin ligase BirA (Fig. 1b, left). Low-level expression reduces the likelihood that non-physiological interactions will be identified or that the stoichiometry of protein complexes will be perturbed¹¹. Biotinylated Nanog (^{bio}Nanog) is efficiently recovered from ES cell nuclear extracts with streptavidin (SA)-agarose (Fig. 1b right), and is biologically active, because its expression attenuates the dependence of ES cells on leukaemia inhibitory factor (LIF; Supplementary Fig. S1). Cells coexpressing BirA and ^{bio}Nanog are functionally equivalent to wild-type ES cells in stem-cell marker gene expression (Supplementary Fig. S1). On size fractionation of ES cell nuclear extracts, ^{bio}Nanog and endogenous Nanog behave similarly, broadly distributed with an apparent molecular mass of more than 160 kDa to 1 MDa, far greater than that of the predicted polypeptide (34 kDa) (Fig. 1c).

To identify Nanog-associated proteins we used single-step capture on streptavidin beads¹⁰, as well as tandem purification with anti-Flag immunoprecipitation followed by capture with streptavidin. Nuclear extracts of ES cells expressing either BirA alone or BirA plus ^{bio}Nanog were processed in parallel. Total peptides sequenced and proteins

identified by whole-lane liquid chromatography–tandem mass spectrometry (LC–MS/MS) are summarized in Fig. 1d (left panel; see also Supplementary Data). With criteria described in the Supplementary Methods, putative Nanog-associated proteins were chosen (Fig. 1d, right panel). The candidates, which are predominantly transcription factors or components of transcriptional complexes, include proteins previously studied in an ES cell context (for example Oct4 or Dax1) and fall into three groups. The first group includes proteins present in at least two of three independent one-step purifications and tandem purification. These proteins (Sall1, Sall4, Rif1, Tiflβ, Mybbp, Dax1 and Nac1) are prime interaction candidates. The second group includes proteins (Zfp281, Err2, Elys, Oct4, Zfp198, NF45 and HDAC2) that may be part of unstable or transient complexes that dissociate during tandem purification. Proteins in the third group (REST, Sp1 and Wapl) seem ‘masked’ in one-step purifications but are recovered in the tandem procedure. Proteins identified by our approach may interact directly with Nanog, or more often indirectly through their association with other components of a protein complex.

We then validated selected Nanog-interacting candidates by performing co-immunoprecipitation experiments. Interactions between Nanog and Dax1, Nac1, Zfp281 and Oct4 were confirmed in heterologous 293T (Supplementary Fig. S2) and ES (Fig. 2a–d) cells transiently transfected with constructs as indicated (see the figure legend for details). In addition, we generated ES cells expressing biotinylated Nac1, Zfp281, Dax1 and Oct4, at or below endogenous protein levels (Supplementary Fig. S3), for affinity capture with anti-streptavidin-agarose and western blotting with anti-streptavidin-horseradish peroxidase (HRP) antibody. We confirmed that Nanog co-purifies with ^{bio}Nac1, ^{bio}Zfp281, ^{bio}Dax1 and ^{bio}Oct4 (Fig. 2e) and that Nac1 co-purifies with ^{bio}Nanog (Fig. 2f). Furthermore, we confirmed the association of Rif1 with endogenous Nanog (Fig. 2g), as well as ectopically expressed Nanog (Fig. 2h, i).

To test the functional relevance of selected components, we employed inhibition of endogenous RNAs by short hairpin RNA (shRNA; Supplementary Fig. S4). If Nanog-associated proteins influence Nanog’s transcriptional activity, their loss of function would be expected to affect ES cell pluripotency. We first examined shRNAs directed to Dax1 and Sall4 in ES cells that express green fluorescent protein (GFP) under the control of Oct4 regulatory elements. In both instances a loss of pluripotency was observed, as revealed by diminished Oct4 expression and concomitant derepression of selected lineage-specific markers (Supplementary Fig. S5). Recent conditional gene targeting of Dax1 (ref. 12) and Sall4 (ref. 13) provides independent support for their crucial function in ES cells.

Further functional validation included shRNA inhibition of two Nanog-associated proteins, Nac1 and Zfp281, not previously

¹Division of Hematology–Oncology, Children’s Hospital and Dana Farber Cancer Institute, Harvard Medical School, Harvard Stem Cell Institute, ²Howard Hughes Medical Institute, Boston, Massachusetts 02115, USA.

examined in ES cells. Nac1 is a BTB-domain containing protein^{14,15} related to *Drosophila* bric-a-brac/tramtrack, which prevents inappropriate neural gene expression. Zfp281, the mouse homologue of human zinc-finger protein ZBP99, may regulate cell proliferation, differentiation and oncogenesis¹⁶. We infected wild-type ES cells with retrovirus containing an shRNA vector carrying a GFP cassette (Supplementary Fig. S4) and isolated GFP^{high} clones harbouring Nac1 and Zfp281 shRNAs. Two clones each with similar extents of RNA knockdown (about 80% for Nac1 and about 70% for Zfp281; Fig. 3a) were studied in detail. Nac1 or Zfp281 knockdown cells maintained on a feeder layer formed small colonies, yet they retained proper ES cell morphology (Fig. 3b, top panels). On passage to gelatin in the presence of LIF, knockdown cells grew poorly and formed disorganized clumps of small cells (Fig. 3b, bottom panels). Striking derepression of primitive endoderm (*Gata6/4*), mesoderm/visceral endoderm (*Bmp2*) and neuroectoderm (*Isl1*) markers was observed, despite normal expression of the stem-cell markers (*Nanog*, *Oct4* and *Rex1*; Fig. 3c). Validation of these findings was achieved through the use of an independent shRNA directed to Zfp281 and by examination of gene-targeted Nac1^{+/-} ES cells (Supplementary Figs S4

and S6). To assess the relevance of Nac1 and Zfp281 with respect to transcriptional control, we examined chromatin occupancy of these proteins at the *Gata6* promoter in wild-type ES cells. As shown in Fig. 3d, Nac1 and Zfp281, as well as Nanog, are highly enriched at the *Gata6* promoter. These data strongly implicate Nac1 and Zfp281, along with Nanog, in the tight control of this crucial target gene and also show that derepression of *Gata6* expression in knockdown cells is not due to off-target effects of shRNAs.

The broad pattern of Nanog on size fractionation, and the identification and validation, both physically and functionally, of associated proteins suggest that Nanog is a component of multiple protein complexes. By performing similar proteomic analyses on several of the Nanog-associated proteins, as well as another stem-cell marker protein Rex1 (ref. 17), we sought to identify additional previously unknown proteins and components shared between complexes and to develop a protein network (Fig. 4a). The size distributions of complexes containing Oct4, Dax1, Nac1, Zfp281 and Rex1 overlap that of Nanog (Supplementary Fig. S7). We employed ES cell lines expressing ^{bio}Dax1, ^{bio}Nac1, ^{bio}Zfp281, ^{bio}Oct4 and ^{bio}Rex1 (Fig. 2e, and Supplementary Fig. S3) for the

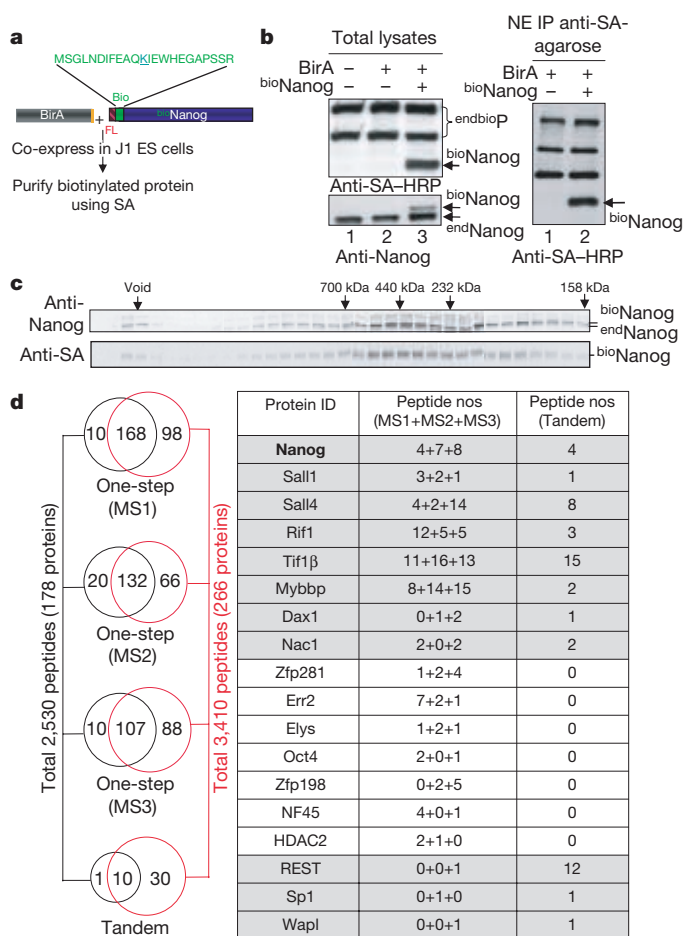


Figure 1 | Protein complexes containing Nanog protein in mouse ES cells. **a**, Strategy for biotinylation of Nanog in ES cells. Flag (FL) and 23-amino-acid biotin (Bio) tags are indicated. SA, streptavidin. **b**, Western blot analyses of expression of biotinylated Nanog (^{bio}Nanog) as well as endogenous Nanog (^{end}Nanog) (left), and binding of ^{bio}Nanog to SA beads (right). IP, immunoprecipitation. **c**, Western blots of fractions using anti-Nanog (top) and anti-SA-HRP antibodies (bottom). **d**, Summary of LC-MS/MS results from three independent one-step purifications and a tandem purification. Total peptides sequenced and proteins identified in BirA (black circles) and ^{bio}Nanog (red circles) samples are indicated. Components of Nanog complexes selected with the criteria described in Supplementary Methods are listed in the table.

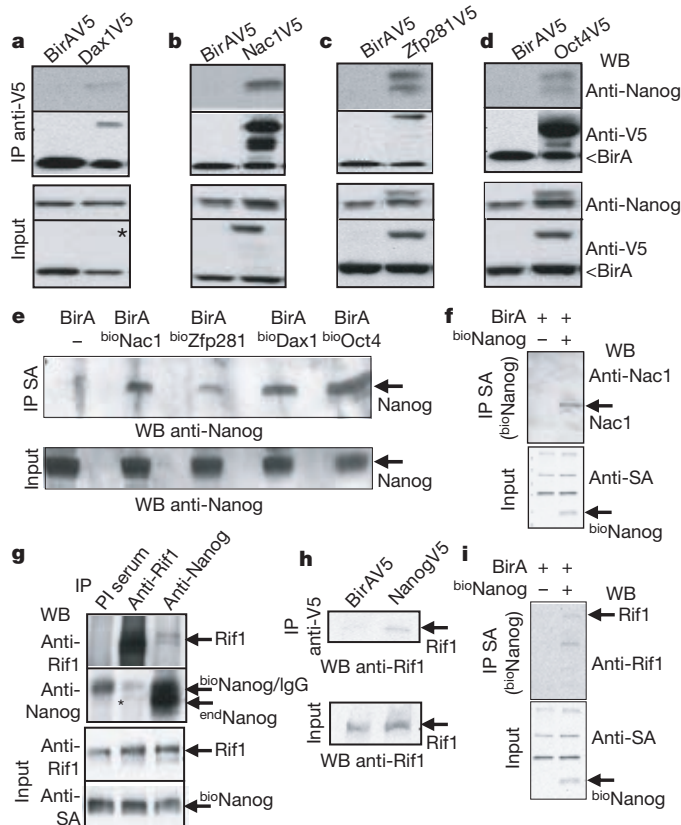


Figure 2 | Confirmation of Nanog association by co-immunoprecipitation in ES cells. **a–d**, Co-immunoprecipitation of Nanog and Dax1 (**a**), Nac1 (**b**), Zfp281 (**c**) and Oct4 (**d**) in ES cells expressing BirA without (**a, b**) or with (**c, d**) ^{bio}Nanog. The asterisk marks a faint band in **a**. IP, immunoprecipitation. **e, f**, Co-immunoprecipitation of Nanog and Nac1, Zfp281, Dax1 and Oct4 in ES cells expressing ^{bio}Nac1, ^{bio}Zfp281, ^{bio}Dax1, ^{bio}Oct4 (**e**) and ^{bio}Nanog (**f**). **g–i**, Co-immunoprecipitation of Nanog and Rif1 in ES cells. In **g**, nuclear extracts from ES cells expressing BirA and ^{bio}Nanog were incubated with preimmune (PI) serum, anti-Rif1 antibody, and anti-Nanog antibody, respectively, followed by capture with protein-A/G agarose. The asterisk indicates endogenous Nanog immunoprecipitated by anti-Rif1 antibody. Note that ^{bio}Nanog migrates together with the IgG band. Antibodies for IP and western blot (WB) analyses are indicated. In **h** and **i**, experiments similar to those in **a** and **f** were performed except that different constructs and antibodies were used.

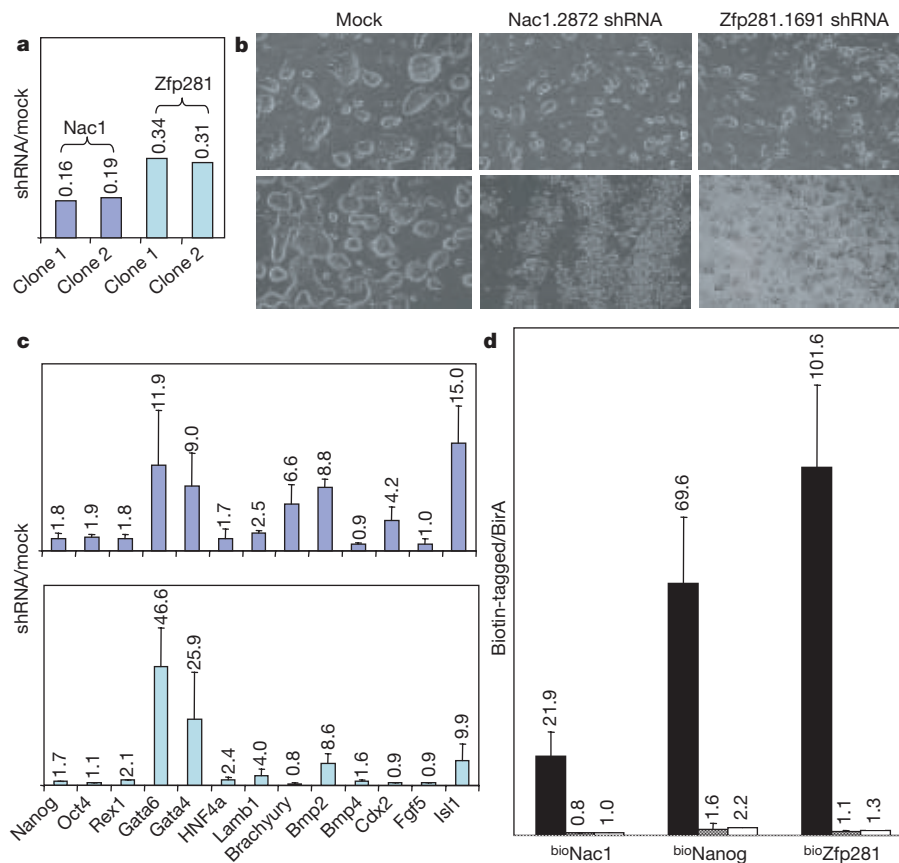


Figure 3 | Functional validation by RNA-mediated interference.

a, Quantitative reverse transcriptase–polymerase chain reaction (RT–PCR) analysis of Nac1 and Zfp281 gene expression (two clones each) after shRNA-mediated knockdown in wild-type ES cells. **b**, Morphology of ES cells infected with virus containing empty shRNA vector (Mock; left), Nac1 shRNA (middle) or Zfp281 shRNA (right). ES cells were grown either on feeders (top) or on gelatin in the presence of LIF (bottom). **c**, Quantitative

RT–PCR analysis of stem-cell markers as well as lineage-specific markers as indicated in Nac1.2872 knockdown cells (top) and Zfp281.1691 knockdown cells (bottom). **d**, Quantitative real-time chromatin immunoprecipitation of biotin-tagged proteins on *Gata6* (black bars) as well as *Gapdh* (grey bars) and β_2 microglobulin (white bars) promoter sequences (as negative controls). Results are shown as means and s.e.m. from two clones in **c** and three independent experiments in **d**.

affinity purification of complexes, as described for ^{bio}Nanog. Proteins were identified with the criteria applied for the selection of Nanog-associated proteins and are summarized in Fig. 4a. Multiple proteins are shared between the different complexes. A protein interaction map (or mini-interactome)¹⁸ summarizes these observations (Fig. 4b). Several features suggest that the network is specialized to establish and/or maintain ES cell pluripotency.

First, the network is remarkably enriched for proteins that have been shown, or are likely, to be required individually for controlling the survival or differentiation of the inner cell mass, or aspects of early development. These include 15 proteins (Fig. 4b, green circles) with known phenotypic alleles in published knockout studies or communicated to us (Supplementary Table S1), and 4 proteins (blue circles) from shRNA-mediated inhibition described previously¹⁹ or in this study (Fig. 3, and Supplementary Figs S5 and S6). Three other proteins (yellow circles) have known phenotypes that are not consistent with an early requirement in development (Supplementary Table S1). Thus, more than 80% of proteins for which knockout or knockdown studies have been performed seem essential for early development and/or ES cell properties. To approximate the 'expected' fraction of essential genes, we curated 300 random transcription factor genes and found that less than 15% of those expressed in ES cells had knockout phenotypes consistent with an early requirement (data not shown).

Second, most genes encoding the proteins within the network are co-regulated, and specifically downregulated, on ES cell differentiation. In Fig. 4c, a rank order depicts the expression of transcripts

for the network proteins relative to Rex1. For comparison, a 'control' set of randomly chosen gene transcripts was similarly analysed (Supplementary Fig. S8). We chose Rex1 as a comparator because it is abruptly and extensively downregulated on the differentiation of ES cells. Co-regulation of genes in the network is consistent with their participation in a common cellular function or pathway.

Third, in reinforcement of this hypothesis, we find that many (at least 56%) of the genes encoding proteins of the mini-interactome have been reported as putative Nanog and/or Oct4 targets in recent ChIP-on-chip²⁰ or ChIP-PET¹⁹ analyses of chromatin occupancy in human or mouse ES cells (Supplementary Table S2). Thus, 'downstream' gene targets of Nanog and Oct4 also serve as 'upstream' effectors in the pluripotency network. Although this feature may stabilize aspects of the ES cell state, it also renders the network vulnerable to attack at multiple points rather than at very few²¹.

Fourth, the network is linked to several cofactor pathways, which are largely involved in mediating transcriptional repression. These include the histone deacetylase NuRD (P66 β and HDAC2), polycomb group (YY1, Rnf2 and Rybp) and SWI/SNF chromatin remodelling (BAF155) complexes. Of potential interest, Rex1 and Oct4, as contrasted with Nanog or its closest partners (for example Dax1, Sall4 and Nac1), are associated with polycomb components. Nanog is linked to HDAC/NuRD through its association with Sall1/4 (ref. 22) and Nac1 (ref. 23). The involvement of multiple co-repressor complexes provides both a method of regulating different sets of target genes and a fail-safe mechanism to prevent differentiation

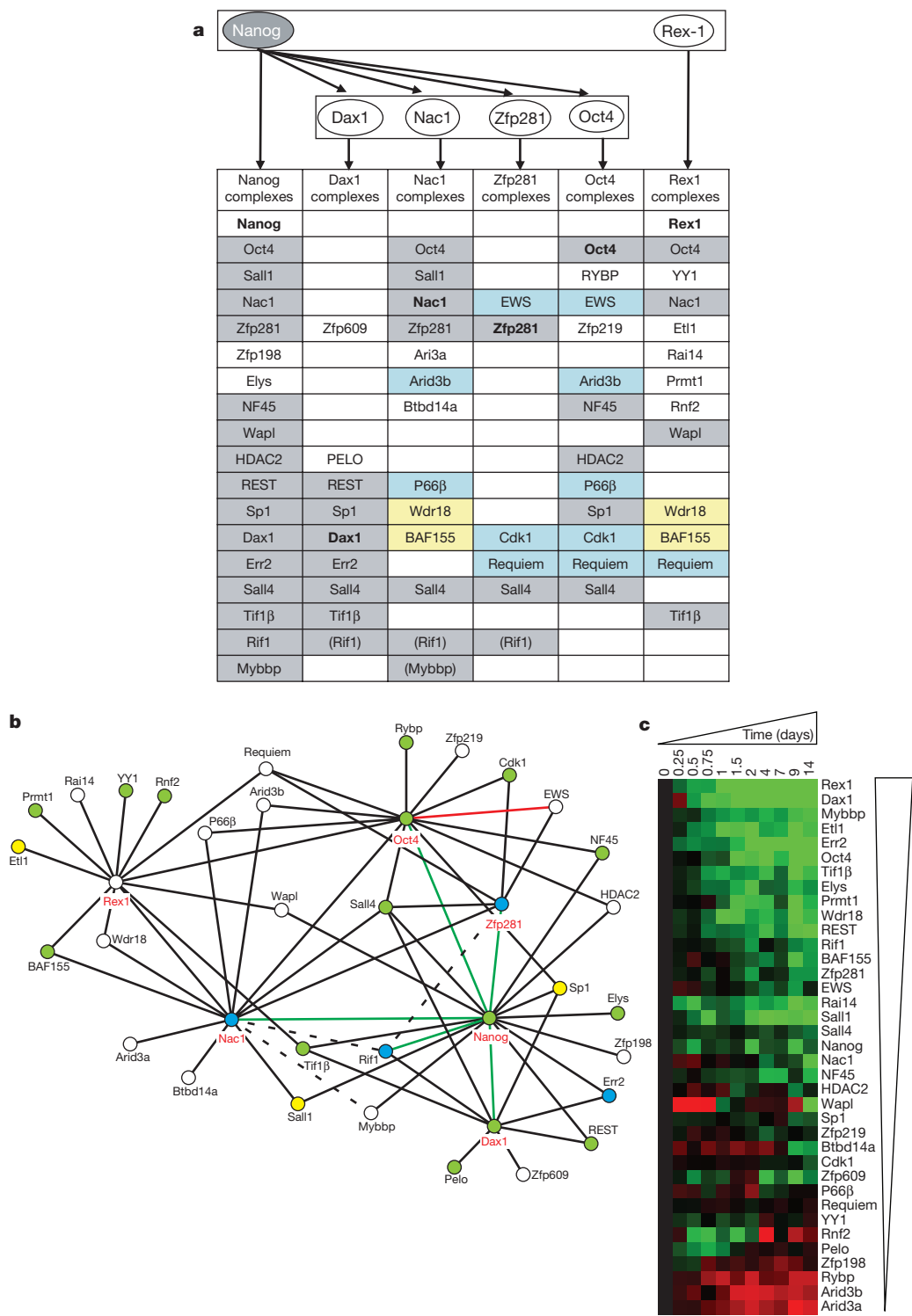


Figure 4 | A protein interaction network in ES cells. **a**, Schematic depiction of the strategy for mapping protein–protein interaction network in ES cells. Starting with the primary tagged baits Nanog and another stem-cell marker protein Rex1, interaction partners of both Nanog and Rex1 were identified. The validated interactors such as Dax1, Nac1, Zfp281 and Oct4 served as secondary tagged baits for the identification of tertiary interacting partners. Selected high-confidence components of multiprotein complexes are listed in the table. Proteins shared by multiple tagged baits are shaded similarly. Proteins in parentheses are present predominantly in tagged samples with some peptide presence in one of the controls during one-step purifications. **b**, Depiction of the features of the interactome. Proteins with red labels are

tagged baits for affinity purification. Green lines indicate interactions confirmed by co-immunoprecipitation and the red line indicates interaction confirmed in the literature²⁷. Green circles indicate proteins whose knockout results in defects in proliferation and/or survival of the inner cell mass or other aspects of early development; blue circles indicate proteins whose reduction by RNA-mediated interference results in defects in self-renewal and/or differentiation of ES cells; yellow circles are proteins whose knockout results in later developmental defects; white circles denote proteins for which no loss-of-function data are available. **c**, Co-regulation of proteins in the network on differentiation. Microarray data were curated from the literature²⁸ and are presented in rank order relative to Rex1.

along different lineages, which is a requisite for maintaining the pluripotent state.

Last, the multiple interactions between critical factors also predict that proteins within the network influence the function of other members, either positively or negatively, and that the stoichiometry within complexes is important for maintaining pluripotency. This latter conclusion is consistent with dosage-dependent effects for Oct4 (ref. 24) and Nanog²⁵. Given the numerous gene targets of the proteins within the network, it is likely that subsets of targets are regulated by different complexes. We have shown that *Gata6* promoter sequences are bound by Nac1, Zfp281 and Nanog. Presumably, proteins within the network, such as Dax1 and Sall4, also bind other targets with Nanog, Oct4 and/or Sox2 (among other proteins). Accordingly, the combinatorial control of target genes will be far greater in complexity than suggested by recent chromatin occupancy studies^{19,20}.

Starting from a central regulator of transcription in ES cells, namely Nanog, followed by iterative tagging and purification of Nanog-associated proteins, we have developed a protein interaction network. Taken together, the extraordinary enrichment for essential genes, the coexpression of most genes, and their roles as both targets and effectors indicate that this mini-interactome might serve as a functional module committed to maintaining ES cell pluripotency. Whereas proteins of the network cooperate to maintain pluripotency, their interdependence renders the cell susceptible to rapid loss of pluripotency on the downregulation, or inactivation, of any one of many components. How differences between mouse and human ES cells are reflected in the network, and whether components of the network that are not strictly specific to ES cells are employed by other types of stem cells for controlling self-renewal versus differentiation, are topics worthy of future study. Last, the network provides a framework for exploring the combinations of factors that may permit faithful reprogramming of differentiated cells to an ES cell state²⁶.

METHODS

In vivo biotinylation, affinity purification and mass spectrometry. J1 ES cells were transfected with a plasmid construct expressing BirAV5his followed by selection with G418 (300 µg ml⁻¹). Clones were picked and expanded for western blotting by using anti-V5-HRP (Invitrogen). A BirA-expressing ES clone was then used for transfection with plasmid constructs containing biotin (or Flag-biotin)-tagged cDNAs. ES cells expressing tagged cDNA were identified by western blotting with anti-streptavidin-HRP of the total lysates or nuclear extracts. One-step affinity purification with streptavidin-agarose was performed as described¹⁰. In brief, nuclear extracts from ES cells expressing BirA and biotin-tagged cDNA were incubated with streptavidin-agarose in a buffer containing 1 × Tris-buffered saline (TBS), 350 mM NaCl and 0.3% Nonidet P40. Binding was performed at 4 °C for 1 h to overnight on a rocking platform, followed by six washes in binding solution. Bound material was eluted by boiling for 5 min in Laemmli buffer, fractionated on a 10% SDS-polyacrylamide gel, and subjected to LC-MS/MS sequencing and data analysis (see Supplementary Methods for details).

For tandem purification, nuclear extracts were incubated overnight with anti-M2 Flag-agarose (Sigma) at 4 °C on a rocking platform. On day 2, after extensive washing with 1 × TBS, bound materials were eluted by using Flag peptide (Sigma). The pooled eluate was then used for streptavidin-agarose capture as described above.

Details of other methods are provided in Supplementary Materials.

Received 10 April; accepted 27 September 2006.

Published online 8 November 2006.

1. Martin, G. R. Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. *Proc. Natl Acad. Sci. USA* **78**, 7634–7638 (1981).
2. Evans, M. J. & Kaufman, M. H. Establishment in culture of pluripotential cells from mouse embryos. *Nature* **292**, 154–156 (1981).
3. Donovan, P. J. & Gearhart, J. The end of the beginning for pluripotent stem cells. *Nature* **414**, 92–97 (2001).
4. Prell, K., Zink, N. & Wolf, E. Pluripotent stem cells—model of embryonic development, tool for gene targeting, and basis of cell therapy. *Anat. Histol. Embryol.* **31**, 169–186 (2002).

5. Silva, J., Chambers, I., Pollard, S. & Smith, A. Nanog promotes transfer of pluripotency after cell fusion. *Nature* **441**, 997–1001 (2006).
6. Mitsui, K. *et al.* The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells. *Cell* **113**, 631–642 (2003).
7. Chambers, I. *et al.* Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells. *Cell* **113**, 643–655 (2003).
8. Nichols, J. *et al.* Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4. *Cell* **95**, 379–391 (1998).
9. Avilion, A. A. *et al.* Multipotent cell lineages in early mouse development depend on SOX2 function. *Genes Dev.* **17**, 126–140 (2003).
10. de Boer, E. *et al.* Efficient biotinylation and single-step purification of tagged transcription factors in mammalian cells and transgenic mice. *Proc. Natl Acad. Sci. USA* **100**, 7480–7485 (2003).
11. Krogan, N. J. *et al.* Global landscape of protein complexes in the yeast *Saccharomyces cerevisiae*. *Nature* **440**, 637–643 (2006).
12. Niakan, K. K. *et al.* Novel role for the orphan nuclear receptor Dax1 in embryogenesis, different from steroidogenesis. *Mol. Genet. Metab.* **88**, 261–271 (2006).
13. Sakaki-Yumoto, M. *et al.* The murine homolog of SALL4, a causative gene in Okihiro syndrome, is essential for embryonic stem cell proliferation, and cooperates with Sall1 in anorectal, heart, brain and kidney development. *Development* **133**, 3005–3013 (2006).
14. Mackler, S. A., Homan, Y. X., Korutla, L., Conti, A. C. & Blendy, J. A. The mouse nac1 gene, encoding a cocaine-regulated Bric-a-brac Tramtrac Broad complex/Pox virus and Zinc finger protein, is regulated by API. *Neuroscience* **121**, 355–361 (2003).
15. Mackler, S. A. *et al.* NAC-1 is a brain POZ/BTB protein that can prevent cocaine-induced sensitization in the rat. *J. Neurosci.* **20**, 6210–6217 (2000).
16. Law, D. J., Du, M., Law, G. L. & Merchant, J. L. ZBP-99 defines a conserved family of transcription factors and regulates ornithine decarboxylase gene expression. *Biochem. Biophys. Res. Commun.* **262**, 113–120 (1999).
17. Thompson, J. R. & Gudas, L. J. Retinoic acid induces parietal endoderm but not primitive endoderm and visceral endoderm differentiation in F9 teratocarcinoma stem cells with a targeted deletion of the Rex-1 (Zfp-42) gene. *Mol. Cell. Endocrinol.* **195**, 119–133 (2002).
18. Batagelj, V. & Mrvar, A. Pajek—program for large network analysis. *Connections* **21**, 47–57 (1998).
19. Loh, Y. H. *et al.* The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. *Nature Genet.* **38**, 431–440 (2006).
20. Boyer, L. A. *et al.* Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell* **122**, 947–956 (2005).
21. Albert, R., Jeong, H. & Barabasi, A. L. Error and attack tolerance of complex networks. *Nature* **406**, 378–382 (2000).
22. Lauberth, S. M. & Rauchman, M. A conserved twelve amino acid motif in sall1 recruits nuRD. *J. Biol. Chem.* **281**, 23922–23931 (2006).
23. Korutla, L., Wang, P. J. & Mackler, S. A. The POZ/BTB protein NAC1 interacts with two different histone deacetylases in neuronal-like cultures. *J. Neurochem.* **94**, 786–793 (2005).
24. Niwa, H., Miyazaki, J. & Smith, A. G. Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. *Nature Genet.* **24**, 372–376 (2000).
25. Hatano, S. Y. *et al.* Pluripotential competence of cells associated with Nanog activity. *Mech. Dev.* **122**, 67–79 (2005).
26. Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**, 663–676 (2006).
27. Lee, J., Rhee, B. K., Bae, G. Y., Han, Y. M. & Kim, J. Stimulation of Oct-4 activity by Ewing's sarcoma protein. *Stem Cells* **23**, 738–751 (2005).
28. Perez-Iratxeta, C. *et al.* Study of stem cell function using microarray experiments. *FEBS Lett.* **579**, 1795–1801 (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank A. Cantor for biotin tagging constructs and for advice on affinity purification and chromatography; Y. Fujiwara and M. Kaku for technical assistance and for Oct4-GFP ES cells; T. De Lange for R1f1 antibody; S. Mackler for Nac1 antibody; S. Lowe and S. Elledge for shRNA vectors; M. Vidal and A.-L. Barabasi for advice and discussion on networks; and R. Tomaino and S. Gygi for performing LC-MS/MS and for providing advice and assistance in data collection and analysis. S.H.O. is an Investigator of the HHMI.

Author Contributions S.R. and J.C. contributed equally to this study. J.W. and S.H.O. conceived and initiated the study. J.W., J.C., X.S., D.N.L. and T.W.T. performed the experiments. S.R. and J.W. analysed data and bioinformatics of the network. J.W. and S.H.O. wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to S.H.O. (stuart_orkin@dfci.harvard.edu).

Boc is a receptor for sonic hedgehog in the guidance of commissural axons

Ami Okada^{1*†}, Frédéric Charron^{1,2,3*}, Steves Morin², David S. Shin¹, Karen Wong⁴, Pierre J. Fabre^{2,3}, Marc Tessier-Lavigne⁴ & Susan K. McConnell¹

In the spinal cord, sonic hedgehog (Shh) is secreted by the floor plate to control the generation of distinct classes of ventral neurons along the dorsoventral axis¹. Genetic and *in vitro* studies have shown that Shh also later acts as a midline-derived chemoattractant for commissural axons². However, the receptor(s) responsible for Shh attraction remain unknown. Here we show that two Robo-related proteins, Boc and Cdon, bind specifically to Shh and are therefore candidate receptors for the action of Shh as an axon guidance ligand. Boc is expressed by commissural neurons, and targeted disruption of *Boc* in mouse results in the misguidance of commissural axons towards the floor plate. RNA-interference-mediated knockdown of Boc impairs the ability of rat commissural axons to turn towards an ectopic source of Shh *in vitro*. Taken together, these data suggest that Boc is essential as a receptor for Shh in commissural axon guidance.

Growing axons respond to attractive and repulsive guidance cues to navigate to their targets^{3–5}. During spinal cord development, commissural neurons in the dorsal neural tube send axons towards then across the floor plate at the ventral midline, forming axon commissures (Supplementary Fig. 1). Commissural axons project towards the midline in part because they are attracted by netrin-1, a long-range chemoattractant secreted by the floor plate^{6–8}. *In vivo* and *in vitro* analyses of mice and explants mutant for *Netrin-1* indicated that the floor plate secretes an additional diffusible attractant for commissural axons⁷. We previously showed that Shh mimics the floor plate's additional chemoattractant activity and acts through Smo to attract commissural axons to the ventral midline². However, Smo does not bind Shh⁹, and the binding receptor(s) acting with Smo to mediate the effect of Shh in axon guidance remain unknown.

Cdon (cell-adhesion-molecule-related/downregulated by oncogenes) and Boc (biregional Cdon-binding protein) are type I transmembrane orphan receptors consisting of four or five immunoglobulin and two or three fibronectin type III (FNIII) repeats in the extracellular domain, and an intracellular domain with no identifiable motifs (Supplementary Fig. 1a); they have been implicated in enhancing muscle differentiation^{10,11}. This domain architecture is closely related to that of axon guidance receptors of the Robo and DCC (deleted in colorectal cancer) families³. Both Cdon and Boc share a high degree of homology in their extracellular domains and are expressed during early stages of development of the central nervous system^{12,13}. In zebrafish forebrain, the morpholino-mediated knockdown of Boc resulted in axon guidance defects¹⁴, although how Boc participates in axon guidance was not explained. Mice with homozygous mutations in Cdon display a microform of holoprosencephaly (HPE)¹⁵, a developmental defect of the forebrain and midface caused by a failure to delineate the midline^{16,17}. In both humans and

mice, many forms of HPE are caused by disruptions in the Shh signalling pathway, indicating that Cdon and Boc might regulate Shh signalling, a possibility supported by the finding that a Cdon/Boc homologue in *Drosophila*, CG9211/iHog, is required for hedgehog (Hh) signalling¹⁸ and by recent reports demonstrating that Cdon, Boc and iHog bind and signal through the Hh pathway^{19–21}. Taken together, these observations indicated that Cdon and/or Boc might function in Shh-mediated axon guidance in vertebrates.

To explore their involvement in commissural axon guidance, we first examined the expression of Cdon and Boc in the mouse spinal cord during development when axons project to the floor plate. We examined RNA expression by *in situ* hybridization and reporter-gene expression in *Boc* and *Cdon* mutant mice generated by targeting these loci with a genetrap cassette encoding a β -galactosidase–neomycin fusion (β -geo) and human placental alkaline phosphatase (PLAP) reporter genes with the use of the 'targeted trapping' method²² (Supplementary Figs 1 and 2). *In situ* hybridization on spinal cord sections from embryonic day (E) 11.5 showed that *Boc* and *Cdon* are expressed in domains consistent with their expression in the progenitors of commissural neurons (Fig. 1a, b, and Supplementary Fig. 2a). During this period, *Cdon* is expressed exclusively in the dorsal ventricular zone (VZ), a domain composed of mitotically active progenitors (Supplementary Figs 1c and 2a, b). In comparison with *Cdon* expression, the *Boc* expression domain extends more ventrally and laterally, and encompasses the region where postmitotic commissural interneurons initiate their differentiation, as demonstrated by overlap of the sites of expression of *Boc* and *Robo3/Rig-1*, a gene expressed by commissural neurons and known to be essential in commissural axon guidance²³ (Fig. 1a, and Supplementary Fig. 2a). In the case of *Boc* (but not *Cdon*), the PLAP reporter also labelled axons projecting to the floor plate, which seemed to be commissural axons as assessed by co-expression of the surface glycoprotein TAG1 (Fig. 1b, and Supplementary Fig. 2e). This result supports the idea that *Boc* mRNA is expressed in differentiating commissural neurons. However, *Boc* expression seems transient, as demonstrated by its narrow overlap with the domain of *Robo3/Rig1* expression (Fig. 1a). To test more directly whether Boc protein is expressed by commissural axons, we generated antibodies against Boc. Although these antibodies did not work on tissue sections, they revealed Boc immunoreactivity on the cell bodies, axons and growth cones of dissociated commissural neurons (but not most hippocampal neurons, which do not express *Boc*) in culture (Fig. 1c, and data not shown). Taken together, these data suggest that *Boc* is expressed by differentiating commissural neurons and Boc protein is expressed by commissural axons during their growth to the midline.

¹Department of Biological Sciences, Stanford University, Stanford, California 94305, USA. ²Institut de recherches cliniques de Montréal (IRCM), 110 avenue des Pins Ouest, Montréal, Québec H2W 1R7, Canada. ³Department of Medicine, University of Montreal, Montreal, Québec, Canada. ⁴Genentech, Inc., 1 DNA Way, South San Francisco, California 94080, USA. [†]Present address: Department of Neurology, Stanford University School of Medicine, Stanford, California 94305, USA.

*These authors contributed equally to this work.

To assess the functions of these receptors in commissural axon guidance, sections of spinal cord from *Boc* and *Cdon* homozygous mutant E11.5 embryos were stained with antibodies against TAG1. Although commissural axon projections seemed normal in *Cdon*^{+/-} and *Cdon*^{-/-} animals, abnormal projections of TAG1⁺ commissural axons, which also expressed PLAP, were observed in *Boc* mutants (Fig. 2, and Supplementary Fig. 2f). The axons were highly dispersed and invaded the ventral spinal cord with ectopic projections extending over the motor columns (Fig. 2a, d). This phenotype is similar to that observed in mice after conditional removal of the Shh signalling mediator *Smo* in commissural neurons, using a Wnt-1 promoter to drive Cre expression². At E11.5, *Boc*^{-/-} mice seemed morphologically normal; neural tube patterning, including the specification of dorsal cell types, also seemed normal in *Boc* mutants, as assessed by the expression of multiple neural tube markers (Supplementary Fig. 3). These data indicate that *Boc* might act in the same pathway as *Smo* to guide commissural axons in response to Shh.

To test whether *Boc* functions as a receptor for Shh in commissural axon guidance, we expressed a *Boc*-green fluorescent protein (GFP) fusion protein, *Cdon*-GFP fusion protein and a variety of control proteins in COS cells, assaying them for binding to Shh. Only cells expressing *Hhip*, *Cdon* or *Boc*, but not surrounding cells or cells transfected with controls, showed specific binding to Shh (Fig. 3a). Quantitative assessment of the affinity of these interactions with the use of Shh-AP, a fusion protein of Shh with alkaline phosphatase, revealed a dissociation constant (*K*_d) similar to that between netrin-1 and its receptor DCC²⁴ (Fig. 3b, c). The interaction between Shh and

Boc or *Cdon* is direct, because purified N-Shh bound to the affinity-purified ectodomains of *Boc* and *Cdon* with the same affinity as to COS cells expressing *Boc* or *Cdon* (*Boc*-ecto-Fc; Fig. 3c, d). Deletion analysis suggested that the Shh-binding activity of *Boc* was primarily attributable to the third FNIII domain (FNIIIc; Supplementary Fig. 4), and neither *Boc* nor *Cdon* bound to Slit in either assay (data not shown). Together, these results show that *Boc* and *Cdon* bind specifically and directly to Shh, which is consistent with recent reports^{19–21}.

Although defects in *Boc* mutant mice are consistent with *Boc*'s being required for Shh-mediated axon guidance, they are also potentially consistent with an involvement in mediating responses to netrin-1. We therefore examined the ability of *Boc*-deficient neurons to respond to netrin-1 *in vitro*⁸. Netrin-1 stimulated similar levels of axon outgrowth in *Boc* heterozygous and mutant explants (Supplementary Fig. 5), indicating that *Boc* is not required for the growth-promoting activity of netrin-1.

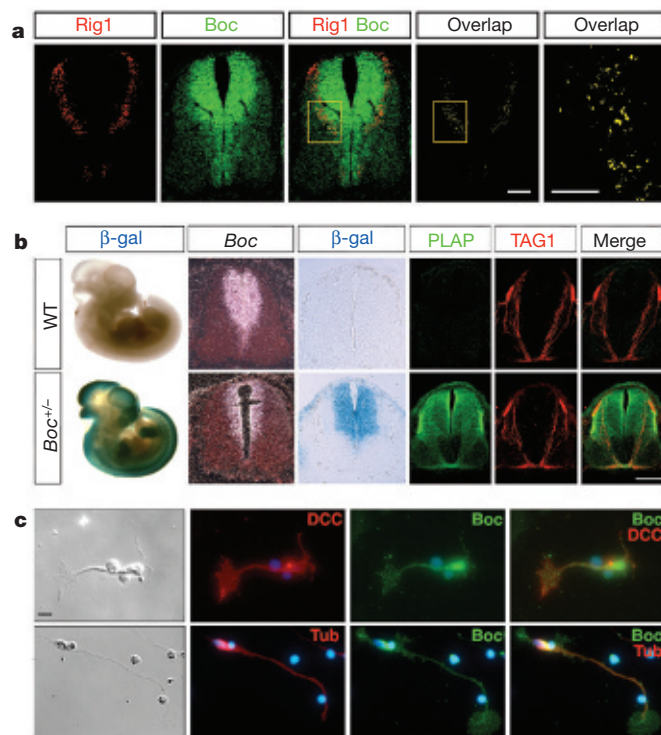


Figure 1 | *Boc* is expressed by young commissural neurons undergoing axon guidance decisions in the spinal cord. **a**, Fluorescent antisense probes to *Robo3/Rig1* (red) and *Boc* (green) on E11.5 spinal cord detect expression overlap, indicating that *Boc* might be expressed by commissural neurons. **b**, Expression of *Boc*, β -gal, PLAP and TAG1 in spinal cord of E11.5 wild-type (WT) and *Boc*^{+/-} embryos. *Boc*-promoter driven PLAP (green) is expressed by dorsal progenitors and by TAG1⁺ commissural neuron axons (red) in *Boc*^{+/-} mice. **c**, *Boc* (green) was detected on the cell body, axon and growth cone of DCC⁺ commissural neurons (red) in dissociated dorsal spinal cord cultures. Acetylated tubulin (Tub; red) labels cell bodies and axons. Scale bars: 100 μ m (**a**, left panels), 50 μ m (**a**, right panel), 200 μ m (**b**) and 20 μ m (**c**).

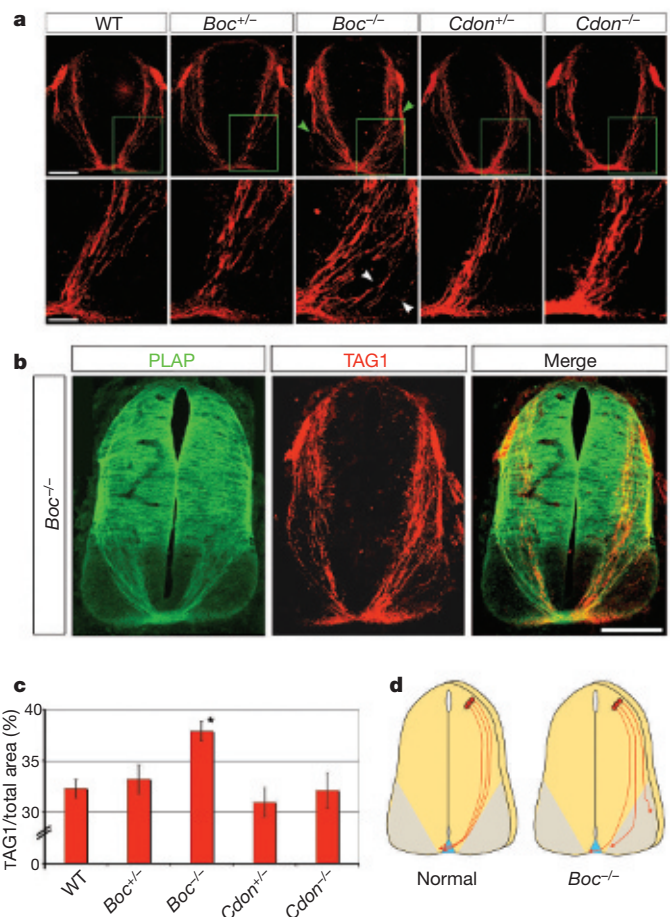


Figure 2 | Aberrant axon targeting in *Boc*^{-/-} but not *Cdon*^{-/-} spinal cord. **a**, Aberrantly projecting TAG1⁺ commissural axon bundles (green arrowheads) and laterally displaced axons invading motor columns (white arrowheads) in E11.5 *Boc*^{-/-} spinal cord; axon projections in *Boc*^{+/-}, *Cdon*^{+/-} and *Cdon*^{-/-} animals seem normal. **b**, *Boc*-promoter driven PLAP is expressed in misguided TAG1⁺ axons in *Boc*^{-/-} embryos (magnification in Supplementary Fig. 2f). **c**, Axon projection deviation was determined by using methods detailed in Supplementary Methods². Number of embryos analysed: wild type (WT), 10 sections (3 embryos); *Boc*^{+/-}, 21 (7); *Boc*^{-/-}, 22 (8); *Cdon*^{+/-}, 13 (4); *Cdon*^{-/-}, 10 (3). Asterisk, *P* < 0.0001 (ANOVA); error bars show standard deviations. **d**, Left: normal commissural axons (red) project from dorsal spinal cord towards the ventral floor plate (blue). Right: in *Boc*^{-/-} mice, axons deviate laterally and invade developing motor columns but seem ultimately to reach the floor plate. Scale bars: 200 μ m (**a**, top), 80 μ m (**a**, bottom) and 200 μ m (**b**).

We next used an *in vitro* commissural axon turning assay to test the function of Boc in Shh-mediated axon turning² (Fig. 4a). After 40 h of co-culture with a source of chemoattractant, commissural axons (which normally project ventrally towards the floor plate) deviate in their trajectories and extend towards the chemoattractant. So far we have successfully used this assay with explants from rat, but not mouse, spinal cord (data not shown); thus, we were unable to examine axon turning directly in *Boc* mutant explants. We therefore designed small interfering RNAs (siRNAs) directed against rat and mouse, but not human, *Boc* (siBoc) to knock down its expression in rat spinal cord neurons. These siRNAs effectively decreased Boc protein levels in COS cells transfected with a mouse (but not a human) *Boc* expression vector and in endogenous rat spinal cord tissue (Supplementary Fig. 6). siBoc was co-electroporated with a GFP expression plasmid into the dorsal region of rat spinal cord explants, which were then co-cultured with COS cells expressing Shh or netrin-1. Axons of commissural neurons electroporated with a control siRNA were attracted towards both Shh and netrin-1 (Fig. 4a, b), as observed in control explants². Commissural axons from explants treated with siRNA against *Boc* showed normal growth and responsiveness to netrin-1 (Fig. 4d), but their ability to turn towards a source of Shh was markedly diminished (Fig. 4b, c). The human *Boc* expression plasmid, which is impervious to the effects of siBoc

in COS cells, successfully rescued the effect of siBoc treatment (Fig. 4c, and Supplementary Fig. 6). Collectively, these results indicate that Boc is required for commissural axons to respond to the chemoattractive effect of Shh.

Our findings show that Boc is expressed in regions of spinal cord harbouring developing commissural neurons, and genetic loss of *Boc* results in commissural axon misguidance without seeming to disrupt spinal cord patterning or the netrin-mediated outgrowth of axons from spinal cord explants. In contrast, loss of *Cdon* does not affect commissural neuron axon guidance, a finding consistent with the more restricted expression of *Cdon* within the dorsal spinal cord to progenitor cells, rather than postmitotic commissural neurons. In addition, our evidence that Boc is expressed by commissural neurons in dissociated culture, that *Boc* promoter-driven PLAP is expressed in the misguided commissural axons, and that Boc is cell-autonomously required for Shh-mediated axon turning are all consistent with the model that Boc is a receptor for Shh during axon guidance.

An earlier study showed defects in axon guidance in zebrafish treated with morpholinos directed against Boc, but the results were interpreted to reflect a role for Boc as a repulsive ligand¹⁴. Our results argue that in the mammalian spinal cord, Boc functions as a receptor for Shh-mediated attraction; whether it can also function in repulsion in other contexts in mammals remains to be explored. Our

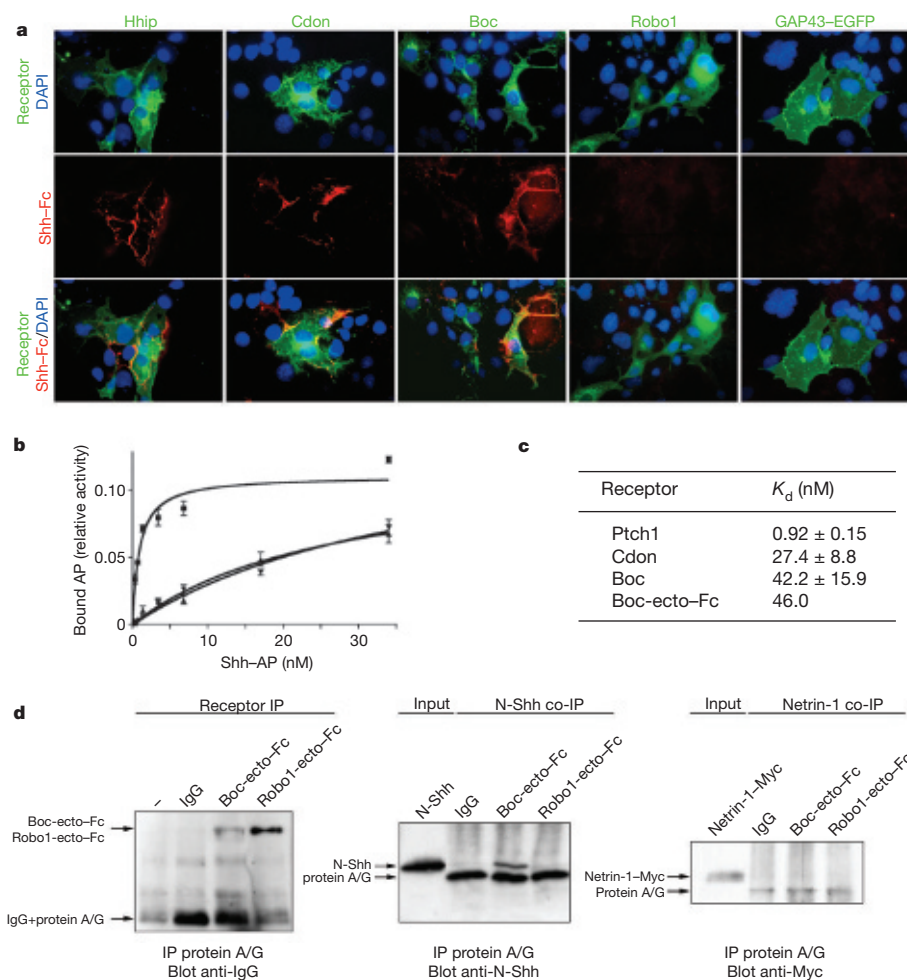


Figure 3 | Boc and Cdon bind directly to Shh. **a**, COS cells expressing Hhip, Boc and Cdon, but not Robo1 or membrane-localized GFP (GAP43–EGFP), bind Fc-epitope-tagged Shh (Shh–Fc; red). **b**, **c**, Quantification of Shh binding affinity to Boc and Cdon. Scatchard analysis was performed on Ptch1-expressing (squares in **b**), Boc-expressing (downward triangles) and Cdon-expressing (upward triangles) cells incubated with a Shh alkaline

phosphatase (Shh–AP) fusion protein to determine dissociation constants (K_d ; $n = 4$) (**c**). Errors are s.e.m. **d**, Boc interacts directly with Shh. The affinity-purified ectodomain of Boc epitope-tagged with Fc (Boc–ecto–Fc; left) co-immunoprecipitated with purified N–Shh, whereas the ectodomain of Robo1 did not (Robo1–ecto–Fc; middle). Purified netrin-1 did not interact with Boc–ecto–Fc or Robo1–ecto–Fc (right). IP, immunoprecipitation.

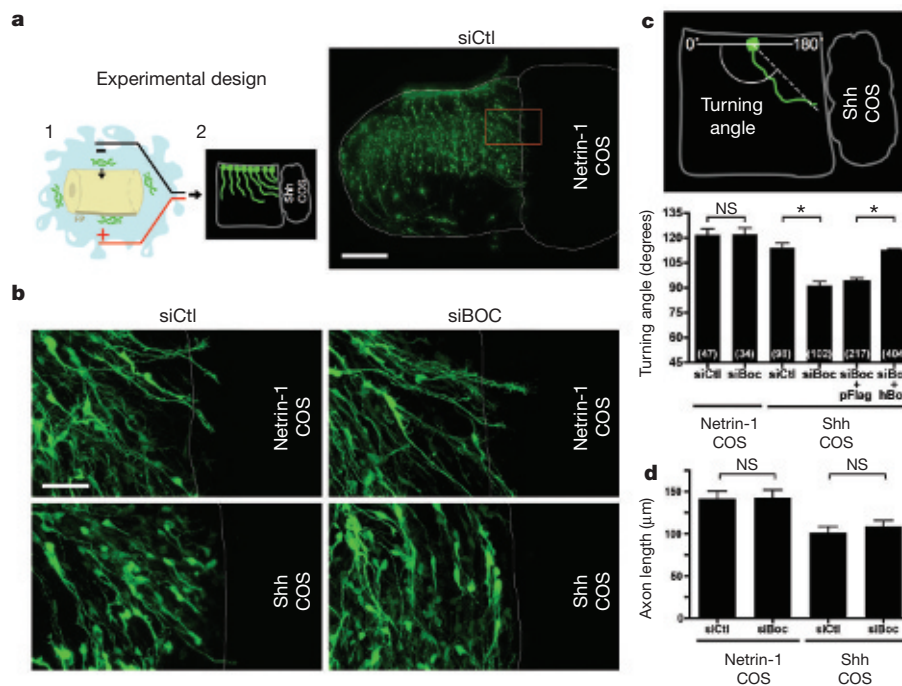


Figure 4 | Boc is required for commissural axons to respond to the chemoattractive effect of Shh. **a**, Left: siRNAs against *Boc* (siBoc) or an unrelated sequence (siCtl) were co-electroporated with GFP into the dorsal region of spinal cord explants (1) and co-cultured with COS cells expressing Shh or netrin-1 (2). Right: axons of commissural neurons co-electroporated with GFP (green) and siCtl are attracted towards netrin-1. Scale bar, 200 μm. **b**, High magnification. Explants electroporated with siBoc showed normal responsiveness to netrin-1 but exhibit markedly diminished ability to turn towards Shh. Scale bar, 50 μm. **c**, Quantification and rescue of siBoc-mediated inhibition of Shh attraction. Top: method of quantification of

axon turning angle. Bottom: siBoc decreases axonal attraction by Shh; co-electroporation of human *Boc* (hBoc, which is impervious to the effects of siBoc), but not control plasmid (pFlag), rescues Shh-mediated attraction. NS, not significant; asterisk, $P < 0.001$. The numbers in parentheses show the number of axons quantified (three to ten explants each; error bars show s.e.m.). **d**, siBoc does not affect axon length. A significant difference in axon length is observed between netrin-1 and Shh ($P < 0.01$), but not between siCtl and siBoc. For siCtl/netrin 21 axons were quantified, for siBoc/netrin 22, for siCtl/Shh 20, and for siBoc/netrin 26.

previous findings established that Shh is a chemoattractant for commissural axons and that its rapid effects are mediated by Smo, as these are inhibited by cyclopamine². Smo itself does not bind to Shh, and our findings suggest a model in which *Boc* is expressed by young commissural neurons, directly binds Shh, and enables the axons of these neurons to respond to Shh, presumably by interacting with Smo either directly or indirectly. However, commissural neurons also express *Ptch* (data not shown), raising the possibility that *Boc* might act in collaboration with other Shh receptors during axon guidance. Although the role of Shh in cell fate specification clearly involves the regulation of transcription in the cell nucleus, axon guidance events at the growth cone are largely independent of transcriptional events⁵. These observations raise the possibility that other, yet to be identified, mechanisms for Shh signalling are deployed downstream of *Boc* and Smo during growth cone guidance.

METHODS

See Supplementary Information for detailed methods.

In situ hybridization. ³⁵S and fluorescent *in situ* hybridization was performed with previously described methods^{23,25}. Template for RNA probes to the cytoplasmic and 3' untranslated regions of murine *Boc* and *Cdon* were generated from complementary DNA isolated from embryonic stem (ES) cells or cDNA obtained from the I.M.A.G.E consortium with the use of standard polymerase chain reaction (PCR) methods.

Generation and analysis of mutant mice. ES cells with mutations in the *Boc* or *Cdon* loci were generated by targeting a gene trap vector 5' to the exon encoding the transmembrane domain, as described previously²². Chimaeric mice were generated from the targeted ES cells, and offspring transmitting the mutant allele were crossed three or more generations into CD1 background.

Binding assays. The binding of Shh to a variety of receptor proteins was examined with published methods^{26–28}. Full-length *Boc* and *Cdon* cDNAs used in the binding experiments were obtained and modified from the I.M.A.G.E.

consortium. *Ptch1* was obtained from C. C. Hui, Shh-AP from A. McMahon, and Shh-Fc from Genentech. Robo1, Robo1-ecto-Fc and gap-EGFP were described previously^{18,29}.

Analysis of axonal projections. Axonal projections were detected by immunohistochemistry and enzymatic reactions as described previously³⁰. Quantification of circumferential misguidance in commissural axons was performed by deriving the ratio of the area of spinal cord spanned by the TAG1-labelled axons to the total area of the spinal cord². Analysis of variance and a two-tailed Student's *t*-test were used to perform statistical analysis.

Turning assays. Rat E11 spinal cord explants were dissected as described previously², then placed in L15 medium containing a β-actin-GFP plasmid and siRNA. The explants were positioned between the electrodes to target the dorsal side preferentially, then electroporated. Explants were subsequently embedded within a collagen matrix, cultured, and analysed as described previously².

Received 11 April; accepted 8 September 2006.

Published online 1 November 2006.

- Ericson, J., Muhr, J., Jessell, T. M. & Edlund, T. Sonic hedgehog: a common signal for ventral patterning along the rostrocaudal axis of the neural tube. *Int. J. Dev. Biol.* **39**, 809–816 (1995).
- Charron, F., Stein, E., Jeong, J., McMahon, A. P. & Tessier-Lavigne, M. The morphogen sonic hedgehog is an axonal chemoattractant that collaborates with netrin-1 in midline axon guidance. *Cell* **113**, 11–23 (2003).
- Tessier-Lavigne, M. & Goodman, C. S. The molecular biology of axon guidance. *Science* **274**, 1123–1133 (1996).
- Charron, F. & Tessier-Lavigne, M. Novel brain wiring functions for classical morphogens: a role as graded positional cues in axon guidance. *Development* **132**, 2251–2262 (2005).
- Dickson, B. J. Molecular mechanisms of axon guidance. *Science* **298**, 1959–1964 (2002).
- Kennedy, T. E., Serafini, T., de la Torre, J. R. & Tessier-Lavigne, M. Netrins are diffusible chemotropic factors for commissural axons in the embryonic spinal cord. *Cell* **78**, 425–435 (1994).
- Serafini, T. *et al.* Netrin-1 is required for commissural axon guidance in the developing vertebrate nervous system. *Cell* **87**, 1001–1014 (1996).

8. Serafini, T. *et al.* The netrins define a family of axon outgrowth-promoting proteins homologous to *C. elegans* UNC-6. *Cell* **78**, 409–424 (1994).
9. Stone, D. M. *et al.* The tumour-suppressor gene patched encodes a candidate receptor for Sonic hedgehog. *Nature* **384**, 129–134 (1996).
10. Kang, J. S., Mulieri, P. J., Hu, Y., Taliana, L. & Krauss, R. S. BOC, an Ig superfamily member, associates with CDO to positively regulate myogenic differentiation. *EMBO J.* **21**, 114–124 (2002).
11. Kang, J. S., Mulieri, P. J., Miller, C., Sassoon, D. A. & Krauss, R. S. CDO, a robo-related cell surface protein that mediates myogenic differentiation. *J. Cell Biol.* **143**, 403–413 (1998).
12. Mulieri, P. J., Okada, A., Sassoon, D. A., McConnell, S. K. & Krauss, R. S. Developmental expression pattern of the cdo gene. *Dev. Dyn.* **219**, 40–49 (2000).
13. Mulieri, P. J., Kang, J. S., Sassoon, D. A. & Krauss, R. S. Expression of the boc gene during murine embryogenesis. *Dev. Dyn.* **223**, 379–388 (2002).
14. Connor, R. M., Allen, C. L., Devine, C. A., Claxton, C. & Key, B. BOC, brother of CDO, is a dorsoventral axon-guidance molecule in the embryonic vertebrate brain. *J. Comp. Neurol.* **485**, 32–42 (2005).
15. Cole, F. & Krauss, R. S. Microform holoprosencephaly in mice that lack the Ig superfamily member Cdon. *Curr. Biol.* **13**, 411–415 (2003).
16. Ming, J. E. & Muenke, M. Holoprosencephaly: from Homer to Hedgehog. *Clin. Genet.* **53**, 155–163 (1998).
17. Muenke, M. & Beachy, P. A. Genetics of ventral forebrain development and holoprosencephaly. *Curr. Opin. Genet. Dev.* **10**, 262–269 (2000).
18. Lum, L. *et al.* Identification of Hedgehog pathway components by RNAi in *Drosophila* cultured cells. *Science* **299**, 2039–2045 (2003).
19. Yao, S., Lum, L. & Beachy, P. The ihog cell-surface proteins bind Hedgehog and mediate pathway activation. *Cell* **125**, 343–357 (2006).
20. Zhang, W., Kang, J. S., Cole, F., Yi, M. J. & Krauss, R. S. Cdo functions at multiple points in the Sonic Hedgehog pathway, and Cdo-deficient mice accurately model human holoprosencephaly. *Dev. Cell* **10**, 657–665 (2006).
21. Tenzen, T. *et al.* The cell surface membrane proteins Cdo and Boc are components and targets of the Hedgehog signaling pathway and feedback network in mice. *Dev. Cell* **10**, 647–656 (2006).
22. Friedel, R. H. *et al.* Gene targeting using a promoterless gene trap vector ('targeted trapping') is an efficient method to mutate a large fraction of genes. *Proc. Natl Acad. Sci. USA* **102**, 13188–13193 (2005).
23. Sabatier, C. *et al.* The divergent Robo family protein rig-1/Robo3 is a negative regulator of slit responsiveness required for midline crossing by commissural axons. *Cell* **117**, 157–169 (2004).
24. Keino-Masu, K. *et al.* Deleted in Colorectal Cancer (DCC) encodes a netrin receptor. *Cell* **87**, 175–185 (1996).
25. Long, H. *et al.* Conserved roles for Slit and Robo proteins in midline commissural axon guidance. *Neuron* **42**, 213–223 (2004).
26. Marigo, V., Davey, R. A., Zuo, Y., Cunningham, J. M. & Tabin, C. J. Biochemical evidence that patched is the Hedgehog receptor. *Nature* **384**, 176–179 (1996).
27. Chuang, P. T. & McMahon, A. P. Vertebrate Hedgehog signalling modulated by induction of a Hedgehog-binding protein. *Nature* **397**, 617–621 (1999).
28. Cheng, H. J. & Flanagan, J. G. Cloning and characterization of RTK ligands using receptor-alkaline phosphatase fusion proteins. *Methods Mol. Biol.* **124**, 313–334 (2001).
29. Okada, A., Lansford, R., Weimann, J. M., Fraser, S. E. & McConnell, S. K. Imaging cells in the developing nervous system with retrovirus expressing modified green fluorescent protein. *Exp. Neurol.* **156**, 394–406 (1999).
30. Mitchell, K. J. *et al.* Functional analysis of secreted and transmembrane proteins critical to mouse development. *Nature Genet.* **28**, 241–249 (2001).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank L. Luo, A. O'Reilly, M. Scott and K. Shen for critically reading the manuscript; C. Jolicoeur and H. Rayburn for the generation of chimaeric mice; C. Kaznowski for technical assistance; R. Krauss, J. Zhang, M. Scott, H. Tian and F. de Sauvage for discussions during early stages of the project; and T. Rando for supporting A.O. during the writing of the manuscript. This work was supported by grants from the National Institute of Mental Health and American Cancer Institute to A.O., the National Eye Institute to S.K.M., the National Institute of Mental Health to M.T.L. and S.K.M., and the Arnold and Mabel Beckman Foundation, the Fonds de Recherche en Santé du Québec (FRSQ), the Canadian Institutes of Health Research (CIHR) and the Peter Lougheed Medical Research Foundation to F.C.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to A.O. (amio@stanford.edu).

LETTERS

Influence of the thalamus on spatial visual processing in frontal cortex

Marc A. Sommer^{1,2} & Robert H. Wurtz²

Each of our movements activates our own sensory receptors, and therefore keeping track of self-movement is a necessary part of analysing sensory input. One way in which the brain keeps track of self-movement is by monitoring an internal copy, or corollary discharge, of motor commands^{1–13}. This concept could explain why we perceive a stable visual world despite our frequent quick, or saccadic, eye movements: corollary discharge about each saccade would permit the visual system to ignore saccade-induced visual changes^{6–9}. The critical missing link has been the connection between corollary discharge and visual processing. Here we show that such a link is formed by a corollary discharge from the thalamus that targets the frontal cortex. In the thalamus, neurons in the mediodorsal nucleus relay a corollary discharge of saccades from the midbrain superior colliculus to the cortical frontal eye field^{10–12}. In the frontal eye field, neurons use corollary discharge to shift their visual receptive fields spatially before saccades^{14,15}. We tested the hypothesis that these two components—a pathway for corollary discharge and neurons with shifting receptive fields—form a circuit in which the corollary discharge drives the shift. First we showed that the known spatial and temporal properties of the corollary discharge predict the dynamic changes in spatial visual processing of cortical neurons when saccades are made. Then we moved from this correlation to causation by isolating single cortical neurons and showing that their spatial visual processing is impaired when corollary discharge from the thalamus is interrupted. Thus the visual processing of frontal neurons is spatiotemporally matched with, and functionally dependent on, corollary discharge input from the thalamus. These experiments establish the first link between corollary discharge and visual processing, delineate a brain circuit that is well suited for mediating visual stability, and provide a framework for studying corollary discharge in other sensory systems.

The dominant hypothesis of how we perceive visual stability is that advance warning of saccades is sent to the visual system (Supplementary Fig. S1a, b)^{7–9}. The only known pathway for saccadic corollary discharge innervates the cortical frontal eye field (FEF)^{10–12} (Supplementary Fig. S1c), which contains visually responsive neurons that have been shown^{14,15} to use corollary discharge. Neurons of this type, which were discovered in the parietal cortex¹⁶, alter their spatial visual processing just before a saccade (Fig. 1a). Specifically, they shift their visual receptive field (RF) to a new location, the future field (FF), where the RF will reside after the saccade^{14–18}. Because the neurons sample the same region of visual space before and after the saccade (in the FF and postsaccadic RF, respectively), they provide information about whether a visual scene is stable across saccades.

We trained monkeys to make saccades while a visual probe appeared at the RF or FF during initial fixation or just before a saccade (Fig. 1b; Supplementary Fig. S2). An example neuron had

a visual response in its RF, but not in its FF, during initial fixation (Fig. 1c, left column), but this spatial sensitivity reversed just before a saccade: the FF became responsive and the RF unresponsive (Fig. 1c, middle column). Saccades alone evoked no activity as revealed in trials without a probe (Fig. 1c, right column, bottom row). Activity in the FF thus reflected a sudden, presaccadic onset of visual sensitivity—a shifting RF.

We searched in the FEFs of two monkeys for neurons with shifting RFs. Only neurons identified as belonging to the superior colliculus (SC)–FEF circuit as defined by antidromic and orthodromic stimulation criteria were studied (antidromic and orthodromic neurons

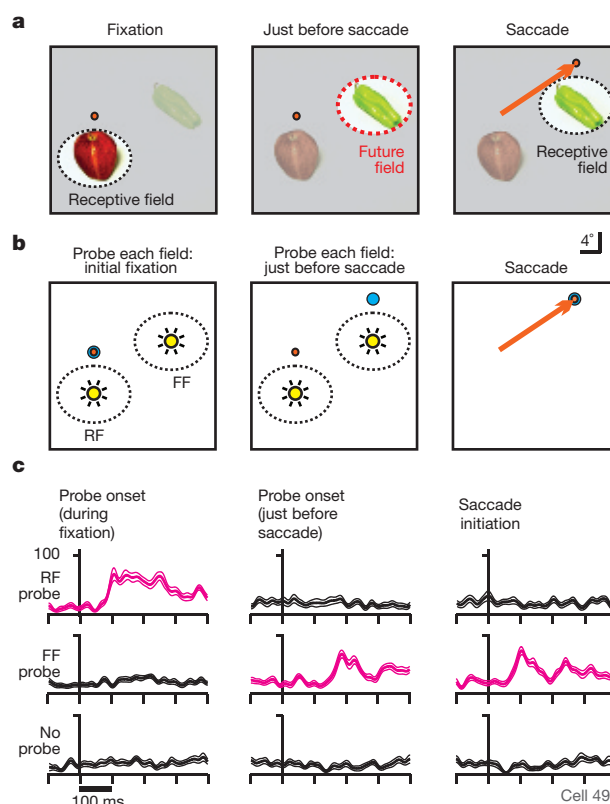


Figure 1 | Shifting receptive fields of visual neurons. **a**, Just before a saccade, visual responsiveness moves from the RF to the FF. **b**, Monkeys looked at (orange dot) a fixation spot (blue dot, left) and made a saccade to a target (blue dot, middle). We flashed a probe (yellow circle) in one of two locations and two times (one probe per trial). **c**, Example FEF neuron. The firing rate (mean $\pm$ s.e.m.; the vertical scale is in spikes s^{-1}) is aligned with events represented in **a** and **b** above. The visual response shifts from the RF (magenta, left) to the FF (magenta, middle) just before the saccade.

¹Department of Neuroscience, the Center for the Neural Basis of Cognition, and the Center for Neuroscience at the University of Pittsburgh, University of Pittsburgh, Pittsburgh, Pennsylvania 15260, USA. ²Laboratory of Sensorimotor Research, National Eye Institute, NIH, Bethesda, Maryland 20892, USA.

behaved identically and were pooled; see Supplementary Notes). Of 71 such neurons that had visual responses, 61% ($n = 43$) had shifting RFs. One-third (14 of 43) showed a complete shift from RF to FF just before saccade initiation (for example Fig. 1c), whereas the rest (29 of 43) became active at the FF while maintaining responsiveness at the RF (Supplementary Fig. S3). The basic properties of the neurons with shifting RFs were similar to those described previously^{14–18}.

If FEF neurons with shifting RFs use corollary discharge from the SC, the shifts should be influenced by the spatiotemporal dynamics of SC activity. Spatially, SC activity encodes saccades on a topographic map¹⁹. Localized activity on the map specifies the vector (eccentricity and direction) of the saccadic target (Fig. 2a, 'hill' of activity). Temporally, SC activity encodes the time of saccade initiation with a volley of action potentials (Fig. 2a, arrow representing sudden onset). Thus we predicted that, spatially, shifting RFs should jump as if directed by a vector specification of the saccade, and temporally, they should move in synchrony with the saccade as if driven by a burst of presaccadic activity.

To study the spatial properties of the shift, we added a visual probe at the midpoint between the RF and the FF. If activity at the midpoint were unchanged before a saccade, this would show that visual sensitivity 'jumped' from the RF to the FF (Fig. 2b, top). However, if activity at the midpoint increased before a saccade, this would refute the prediction of a jump and suggest that a gradual spread of visual

sensitivity travelled from the RF to the FF (Fig. 2b, bottom). An example neuron responded to a probe in the RF during initial fixation (Fig. 2c, top) but had little if any response at the midpoint and was unresponsive at the FF. Just before the saccade, the neuron suddenly became visually responsive at the FF ($P < 0.0001$), whereas the midpoint location showed no significant change ($P > 0.05$; measured 100–300 ms after probe onset). We isolated 13 neurons long enough to test the midpoint, and every one exhibited no change there as activity jumped to the FF. The pooled data from all 13 neurons (Fig. 2c, bottom) showed that just before a saccade, the average activity dropped slightly at the RF, rose markedly at the FF (by 43 spikes s^{-1} , $P < 0.0001$) and did not change significantly at the midpoint ($P > 0.05$). A minor, late increase at the midpoint was due to a slight overlap of the probe onto the edge of the FF (see further analyses in Supplementary Figs S4–S6). Shifting RFs therefore jump, rather than spread, to their new locations.

Next we tested our prediction that shifts were synchronized to saccade initiation by flashing probes earlier to impose a substantial delay (about 200 ms) between probe onset and saccade initiation. If shifts were synchronized to saccade initiation they would start after the delay, when the saccades begin, but if shifts were synchronized to probe onset they would start at a normal visual latency after the probes. On a trial-by-trial basis, we found that the times of shift onset and saccade initiation were tightly correlated in single neurons (Fig. 3a, left; $R = 0.97$, $P < 0.001$) and in the population (Fig. 3b; mean $R = 0.50$, greater than 0 at $P < 0.002$, t -test; $n = 13$ neurons having more than ten trials). In addition, there was a higher peak in the average firing rate when aligned to saccade initiation (Fig. 3a, right) as opposed to probe onset (Fig. 3a, left; true for 24 of 26 individual neurons, an average increase of 6 spikes s^{-1} ; $P = 0.001$, paired t -test). In contrast, shift onset had no clear relation to probe onset; the shift started long after the normal visual latency (green arrow) in the single example (Fig. 3a, left) and in the population (not shown; average shift onset time 174 ± 92 ms (mean $\pm$ s.d.) versus average visual latency 86 ± 16 ms; $P < 0.0001$). In sum, shifting RFs were synchronized with saccades. On average they started with saccade initiation (Fig. 3c; mean 24 ± 74 ms (mean $\pm$ s.d.),

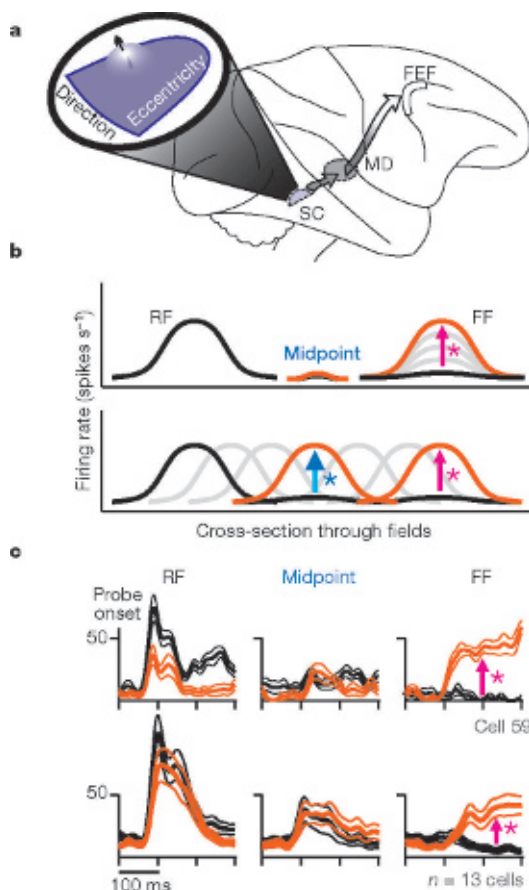


Figure 2 | The spatial properties of shifting RFs are predicted by corollary discharge from the SC–MD–FEF pathway. **a**, The corollary discharge arises from the SC, which encodes saccades spatially, using a map of direction and eccentricity, and temporally, using a burst of activity. **b**, Spatially, we predicted that shifting RFs jump (top), as indicated by a significant (asterisk) increase in activity at the FF (magenta arrow) but no significant difference at the midpoint. Alternatively, shifting RFs could spread (bottom). Probe onset time: black, initial fixation; orange, just before saccade. **c**, Example neuron (top) and population (bottom); data aligned to probe onset. Shifting RFs moved as a jump. The vertical scale is in spikes s^{-1} .

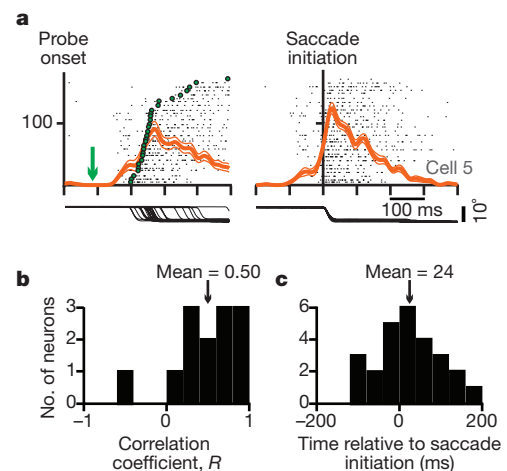


Figure 3 | The temporal properties of shifting RFs are predicted by corollary discharge from the SC–MD–FEF pathway. **a**, Our hypothesis predicts that shifting RFs are synchronized with saccades. Shift activity (in the FF) for an example neuron is aligned to probe onset (left) and saccade initiation (right). Rasters of action potentials from each trial are sorted by saccadic latency (green dots). Eye position traces (horizontal component) are shown below. The green arrow indicates the average visual latency of this neuron. The vertical scale is in spikes s^{-1} . **b**, Strength of correlation (Pearson's R) between shift onset time and saccadic latency for the population. Mean $R > 0$ at $P < 0.002$. **c**, Shift onset time relative to saccade initiation for the population. There is no significant difference in shift onset time from 0 ms.

not significantly different from 0; range -94 to 164 ms). A visual response gated by saccadic initiation would be critical for perceiving visual stability, because saccades can be cancelled about 100 ms before initiation²⁰; only after this 'point of no return', when a saccade is inevitable, should neurons shift their RFs.

Next we looked for evidence that corollary discharge from SC through the thalamus caused the shifting RFs in the FEF. In an experimental session, first we isolated a visually responsive FEF neuron that was connected to the SC and had a shifting RF (as described for the 43 neurons above). While maintaining isolation of the neuron, we inserted an injection syringe needle targeted at the previously identified relay nucleus from the SC to the FEF in the mediodorsal (MD) thalamus (Fig. 4a). We then returned to the FEF neuron, quantified its visual sensitivity at the RF and the FF, and inactivated the MD nucleus with muscimol, an agonist of the inhibitory neurotransmitter γ -aminobutyric acid (GABA)²¹. Finally, we retested the visual sensitivity of the same neuron at the RF and the FF. Our prediction was that MD inactivation should reduce activity at the FF.

We completed the lengthy sequence of procedures required in this experiment for eight FEF neurons (Supplementary Fig. S7). In an example neuron, the major effect was a 70% decrease ($P < 0.0001$) in FF activity (Fig. 4b, lower middle and right), whereas the visual response in the RF was spared (Fig. 4b, upper left), as was low-level

activity in the trials without a probe (not shown). Inactivating the corollary discharge pathway through the MD therefore caused a marked reduction in the ability of this neuron to shift its RF.

Because the SC–MD–FEF pathway on each side of the brain represents only contraversive saccades¹⁰, a further prediction was that our unilateral inactivations would eliminate corollary discharge for contraversive saccades only. Therefore shifts accompanying contraversive saccades, but not ipsiversive saccades, should be impaired. This is exactly what we found (Fig. 4c).

Overall, we found that activity at the FF decreased during inactivation for every neuron (16–70%, individually significant for seven of eight neurons; Supplementary Fig. S7). The average deficit (Fig. 4d, left) was severe in the FF for contraversive saccades (53.9% deficit, $P < 0.001$) but not significant in the RF ($P = 0.48$) or in the FF for ipsiversive saccades ($P = 0.76$). In terms of firing rates (Fig. 4d, right), the average visual response decreased from 69.6 to 30.3 spikes s^{-1} in the FF for contraversive saccades (paired t -test, $P = 0.019$) but was unchanged in the RF ($P = 0.2$) and in the FF for ipsiversive saccades ($P = 0.4$). Data from the trials without a probe never showed a significant deficit (not shown). Most of the neurons were antidromically activated from the SC, but the one orthodromically activated neuron showed the same effect (55% deficit; Supplementary Fig. S7a). Injections as small as $0.4 \mu l$ were effective and larger volumes did not cause larger effects, suggesting that inactivation beyond the targeted MD relay neurons was inconsequential (Table in Supplementary Fig. S7). A side-product of MD inactivation, bilaterally increased saccadic latencies, did not cause the deficits (Supplementary Fig. S8). For more consideration of these issues see the Supplementary Discussion.

We conclude that corollary discharge from the SC–MD–FEF pathway is both appropriate and necessary to produce shifting RFs in FEF neurons. These results delineate a link between an identified corollary discharge and a sensory system in the primate brain. More specifically, the data provide a circuit-level explanation for our percept of visual stability across saccades—an aim of neuroscience for at least 50 years^{1,8,9}. Corollary discharge from the SC causes the visual system to shift its RFs spatially, and these shifting RFs are proposed to promote the percept of visual stability across saccades¹⁶ (although exactly how this would be achieved has not yet been modelled).

Our findings permit the first direct test of the visual-stability hypothesis. In monkeys trained to detect visual motion, one could inactivate MD relay neurons to reduce shifting RFs; if these shifts underlie perceptual stability, the monkeys should report that static visual stimuli move with saccades. More generally, our approach may open up the study of other corollary discharge circuits such as those mediating the percept of a stable soundscape during head movements.

METHODS

Two rhesus monkeys (*Macaca mulatta*), 'T' and 'Y', were prepared for neurophysiological experiments as described in detail elsewhere^{10,22}. Details on physiology, task design and analysis are given in Supplementary Methods. To record from FEF neurons, we used standard extracellular methods combined with antidromic and orthodromic stimulation^{10,22} to confirm a connection with the SC (Supplementary Notes). To inactivate MD relay neurons, we first mapped their location in the thalamus by activating each one antidromically from the FEF and orthodromically from the SC. Relay neurons of the SC–MD–FEF pathway were concentrated in a restricted volume along the lateral edge of the MD thalamus^{10,23}. We inactivated them by inserting the tip of a 30-gauge cannula at their location and injecting $5 \mu g \mu l^{-1}$ muscimol.

Received 29 June; accepted 25 September 2006.

Published online 8 November 2006.

- Teuber, H.-L. In *The Frontal Granular Cortex and Behavior* (eds Warren, J. P. & Akert, K.) 410–444 (McGraw-Hill, New York, 1964).
- Delcomyn, F. Corollary discharge to cockroach giant interneurons. *Nature* **269**, 160–162 (1977).
- Poulet, J. F. & Hedwig, B. A corollary discharge maintains auditory sensitivity during sound production. *Nature* **418**, 872–876 (2002).

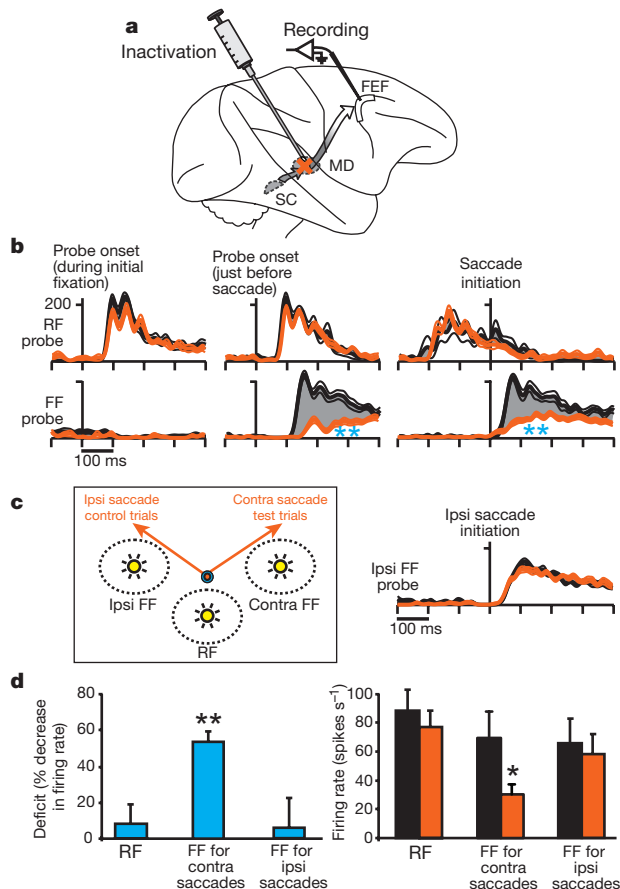


Figure 4 | Shifting RFs require corollary discharge from the pathway. **a**, We recorded from FEF neurons while inactivating MD relay neurons that convey corollary discharge. **b**, Example neuron (the same one as in Supplementary Fig. S3), with the difference in activity between during inactivation (orange) and before inactivation (black) highlighted in grey. Visual responsiveness in the FF just before the saccade—the shift—decreased severely during MD inactivation. The vertical scale is in spikes s^{-1} . **c**, In contrast, in ipsiversive (ipsi) saccade trials for the example neuron (left) there was no deficit (right). Contra, contraversive. **d**, Population data. Left, average percentage deficits; right, average changes in raw visual response (means and s.e.m.). Asterisk, $P < 0.05$; two asterisks, $P < 0.0001$.

4. Bell, C. C. in *Comparative Physiology: Sensory Systems* (eds Bolis, L. & Keynes, R. D.) 637–646 (Cambridge Univ. Press, Cambridge, 1984).
5. Thier, P., Haarmer, T., Chakraborty, S., Lindner, A. & Tikhonov, A. Cortical substrates of perceptual stability during eye movements. *Neuroimage* **14**, S33–S39 (2001).
6. Wurtz, R. H. & Sommer, M. A. Identifying corollary discharges for movement in the primate brain. *Prog. Brain Res.* **144**, 47–60 (2004).
7. Von Helmholtz, H. *Helmholtz's Treatise on Physiological Optics* (Thoemmes, Bristol, 2000).
8. Sperry, R. W. Neural basis of the spontaneous optokinetic response produced by visual inversion. *J. Comp. Physiol. Psychol.* **43**, 482–489 (1950).
9. Von Holst, E. & Mittelstaedt, H. Das Reafferenzprinzip. Wechselwirkungen zwischen Zentralnervensystem und Peripherie. *Naturwissenschaften* **37**, 464–476 (1950).
10. Sommer, M. A. & Wurtz, R. H. What the brain stem tells the frontal cortex. I. Oculomotor signals sent from superior colliculus to frontal eye field via mediodorsal thalamus. *J. Neurophysiol.* **91**, 1381–1402 (2004).
11. Sommer, M. A. & Wurtz, R. H. What the brain stem tells the frontal cortex. II. Role of the SC–MD–FEF pathway in corollary discharge. *J. Neurophysiol.* **91**, 1403–1423 (2004).
12. Sommer, M. A. & Wurtz, R. H. A pathway in primate brain for internal monitoring of movements. *Science* **296**, 1480–1482 (2002).
13. Guthrie, B. L., Porter, J. D. & Sparks, D. L. Corollary discharge provides accurate eye position information to the oculomotor system. *Science* **221**, 1193–1195 (1983).
14. Umeno, M. M. & Goldberg, M. E. Spatial processing in the monkey frontal eye field. I. Predictive visual responses. *J. Neurophysiol.* **78**, 1373–1383 (1997).
15. Umeno, M. M. & Goldberg, M. E. Spatial processing in the monkey frontal eye field. II. Memory responses. *J. Neurophysiol.* **86**, 2344–2352 (2001).
16. Duhamel, J. R., Colby, C. L. & Goldberg, M. E. The updating of the representation of visual space in parietal cortex by intended eye movements. *Science* **255**, 90–92 (1992).
17. Kusunoki, M. & Goldberg, M. E. The time course of perisaccadic receptive field shifts in the lateral intraparietal area of the monkey. *J. Neurophysiol.* **89**, 1519–1527 (2003).
18. Nakamura, K. & Colby, C. L. Updating of the visual representation in monkey striate and extrastriate cortex during saccades. *Proc. Natl Acad. Sci. USA* **99**, 4026–4031 (2002).
19. Sparks, D. L. & Hartwich-Young, R. in *The Neurobiology of Saccadic Eye Movements; Reviews of Oculomotor Research* Vol. III (eds Wurtz, R. H. & Goldberg, M. E.) 213–255 (Elsevier, Amsterdam, 1989).
20. Hanes, D. P. & Schall, J. D. Countermanding saccades in macaque. *Vis. Neurosci.* **12**, 929–937 (1995).
21. Lomber, S. G. The advantages and limitations of permanent or reversible deactivation techniques in the assessment of neural function. *J. Neurosci. Methods* **86**, 109–117 (1999).
22. Sommer, M. A. & Wurtz, R. H. Composition and topographic organization of signals sent from the frontal eye field to the superior colliculus. *J. Neurophysiol.* **83**, 1979–2001 (2000).
23. Lynch, J. C., Hoover, J. E. & Strick, P. L. Input to the primate frontal eye field from the substantia nigra, superior colliculus, and dentate nucleus demonstrated by transneuronal transport. *Exp. Brain Res.* **100**, 181–186 (1994).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This research was supported by the Intramural Research Program of the NIH and the NEI.

Author Contributions Both authors designed the experiments, M.A.S. performed and analysed them, and both authors wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to M.A.S. (sommer@bns.pitt.edu).

LETTERS

Haemagglutinin mutations responsible for the binding of H5N1 influenza A viruses to human-type receptors

Shinya Yamada^{1,3}, Yasuo Suzuki^{3,4}, Takashi Suzuki^{3,5}, Mai Q. Le⁶, Chairul A. Nidom⁷, Yuko Sakai-Tagawa^{1,3}, Yukiko Muramoto^{1,3}, Mutsumi Ito^{1,3}, Maki Kiso^{1,3}, Taisuke Horimoto^{1,3}, Kyoko Shinya⁸, Toshihiko Sawada⁹, Makoto Kiso⁹, Taiichi Usui¹⁰, Takeomi Murata¹⁰, Yipu Lin¹¹, Alan Hay¹¹, Lesley F. Haire¹¹, David J. Stevens¹¹, Rupert J. Russell¹¹†, Steven J. Gamblin¹¹, John J. Skehel¹¹ & Yoshihiro Kawaoka^{1,2,3,12}

H5N1 influenza A viruses have spread to numerous countries in Asia, Europe and Africa, infecting not only large numbers of poultry, but also an increasing number of humans, often with lethal effects^{1,2}. Human and avian influenza A viruses differ in their recognition of host cell receptors: the former preferentially recognize receptors with saccharides terminating in sialic acid- α 2,6-galactose (SA α 2,6Gal), whereas the latter prefer those ending in SA α 2,3Gal (refs 3–6). A conversion from SA α 2,3Gal to SA α 2,6Gal recognition is thought to be one of the changes that must occur before avian influenza viruses can replicate efficiently in humans and acquire the potential to cause a pandemic. By identifying mutations in the receptor-binding haemagglutinin (HA) molecule that would enable avian H5N1 viruses to recognize human-type host cell receptors, it may be possible to predict (and thus to increase preparedness for) the emergence of pandemic viruses. Here we show that some H5N1 viruses isolated from humans can bind to both human and avian receptors, in contrast to those isolated from chickens and ducks, which recognize the avian receptors exclusively. Mutations at positions 182 and 192 independently convert the HAs of H5N1 viruses known to recognize the avian receptor to ones that recognize the human receptor. Analysis of the crystal structure of the HA from an H5N1 virus used in our genetic experiments shows that the locations of these amino acids in the HA molecule are compatible with an effect on receptor binding. The amino acid changes that we identify might serve as molecular markers for assessing the pandemic potential of H5N1 field isolates.

We used H5N1 viruses isolated from birds and humans to identify amino acid changes in the HA molecule that could enable the viruses to recognize human-type receptors (Supplementary Table 1 and Fig. 1). A/Vietnam/30262III/04 and A/Vietnam/3028II/04 contained a heterogeneous mixture of HAs on sequence analysis, prompting us to plaque-purify the viruses in Madin–Darby canine kidney (MDCK) cells to obtain viral clones with distinct HA sequences (Supplementary Table 1). We also used plaque-purified clones of A/Vietnam/30408/05 that differed in their HAs⁷. To expand the repertoire of viruses, we synthesized the HAs of eight H5N1 viruses isolated from humans in Thailand, Vietnam and Cambodia, using sequences in the

Influenza Sequence Database (<https://www.flu.lanl.gov/>). Mutant viruses possessing each of these HA genes, the neuraminidase of A/Vietnam/1194/04 (VN1194) and all remaining genes from A/Puerto Rico/8/34 (PR8; H1N1) were then generated by reverse genetics⁸.

The receptor specificity of the resulting H5N1 viruses was determined with an assay that measured direct binding to sialylglycopolymers possessing either SA α 2,3Gal or SA α 2,6Gal. None of the five avian H5N1 isolates bound appreciably to SA α 2,6Gal, whereas 3 of the 21 human isolates, A/Vietnam/3028II/04clone3 (VN/3028IIcl3), A/Thailand/1-KAN-1/04RG (Thai/KAN), and A/Vietnam/30408/05clone7 (VN/30408cl7), including subclones, bound to both SA α 2,3Gal and SA α 2,6Gal; some of the other human H5N1 viruses also recognized SA α 2,6Gal but to only a limited extent (Fig. 1 and Supplementary Fig. 2). The specificity of the receptor-binding assay was verified as follows. Sialylglycopolymers possessing either SA α 2,3Gal or SA α 2,6Gal were treated with *Arthrobacter ureafaciens* sialidase or mock-treated and then incubated with SA–Gal linkage-specific lectins: *Maackia amurensis* lectin II (MALII), specific for SA α 2,3Gal; *Sambucus nigra* lectin (SNA), specific for SA α 2,6Gal; and *M. amurensis* lectin I (MALI), specific for both Gal β 1-4GlcNAc and SA α 2,3Gal β 1-4GlcNAc. The sialylglycopolymer possessing SA α 2,3Gal reacted only with MALII, whereas that possessing SA α 2,6Gal reacted only with SNA (Supplementary Fig. 3). The sialylglycopolymers treated with the sialidase no longer bound to the respective SA–Gal linkage-specific lectins, but gained reactivity with MALI, confirming that the sialidase treatment removed only the terminal sialic acid. These lectins did not react with polymers lacking oligosaccharides, as expected. Similarly, viruses that bound to the sialylglycopolymers did not bind to those treated with the sialidase or the polymer backbone without oligosaccharides.

To identify mutations capable of conferring SA α 2,6Gal recognition, we focused on the three viruses that bound to the human receptor analogue relatively well: VN/3028IIcl3, Thai/KAN and VN/30408cl7. Comparison of the HA sequences identified two amino acid differences at positions 192 and 223 between VN/3028IIcl3 and VN1194 (a human clade-1 isolate, whose HA is identical to that of avian viruses such as A/duck/Thailand/71.1/2004 and

¹Division of Virology, Department of Microbiology and Immunology, and ²International Research Centre for Infectious Diseases, Institute of Medical Science, University of Tokyo, Tokyo 108-8639, Japan. ³Core Research for Evolutional Science and Technology, Japan Science and Technology Agency, Kawaguchi, Saitama 332-0012, Japan. ⁴College of Life and Health Sciences, Chubu University, Kasugai, Aichi 487-8501, Japan. ⁵Department of Biochemistry, University of Shizuoka, School of Pharmaceutical Sciences and COE Program in the 21st Century, Yada, Shizuoka 422-8526, Japan. ⁶National Institute of Hygiene and Epidemiology, Hanoi, Vietnam. ⁷Avian Influenza Laboratory, Tropical Disease Centre, Airlangga University, Surabaya, Indonesia. ⁸The Avian Zoonosis Research Centre, Tottori University, Tottori 680-8553, Japan. ⁹Department of Applied Bioorganic Chemistry, The United Graduate School of Agricultural Science, Gifu University, Yanagido, Gifu 501-1193, Japan. ¹⁰Department of Applied Biological Chemistry, Shizuoka University, Shizuoka 422-8529, Japan. ¹¹MRC National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, UK. ¹²Department of Pathobiological Sciences, University of Wisconsin-Madison, Madison, Wisconsin 53706, USA. †Present address: Centre for Biomolecular Sciences, University of St Andrews, St Andrews KY16 9ST, UK.

A/chicken/Thailand/9.1/2004 and whose HA structure we determined by X-ray crystallography; see below). Introduction of the Gln192→Arg mutation, but not the Ser223→Asn mutation, into the HA of VN1194 appreciably enhanced the capacity of the HA to recognize SA α 2,6Gal, and introduction of both mutations increased the binding capacity further. This finding implicates Gln192→Arg as a possible determinant of the shift to recognition of the human receptor by VN/3028IIc13 (Fig. 2a). The Thai/KAN virus also showed two amino acid changes in HA1 as compared with VN1194: Gly139→Arg and Asn182→Lys. Introduction of either mutation into the VN1194 HA enhanced SA α 2,6Gal binding (Fig. 2b), and an additional increase in binding capacity was observed when both mutations were substituted simultaneously. Thus, both Gly139→Arg and Asn182→Lys seem to contribute to recognition of the human-type receptor.

The HA of VN/30408cl7 differed from that of VN1194 by four amino acids at positions 75, 123 and 193 in HA1 and 167 in HA2. When introduced singly into the VN1194 HA, none of these amino acid substitutions enhanced binding to SA α 2,6Gal, with the exception of Asn193→Lys, which increased binding capacity slightly (Supplementary Fig. 4a). Different pairs of these amino acid residues substituted at positions 75, 123 and 193 (in HA1) and 167 (in HA2) produced variable increases in SA α 2,6Gal recognition (Supplementary Fig. 4b), and the introduction of various combinations of three mutations into the VN1194 HA enhanced SA α 2,6Gal binding even further (Supplementary Fig. 4c). These results suggest that two

or more of these changes acting in concert are necessary for SA α 2,6Gal recognition by the VN/30408cl7 HA.

A group of H5N1 viruses (clade 2) that are antigenically and genetically distinct from previously circulating viruses (clade 1) have become prevalent¹ (Supplementary Fig. 1). As it was unclear whether mutations that conferred SA α 2,6Gal recognition to the HA of clade-1 VN1194 virus would act comparably in the HAs of viruses of different clades, we inserted the Gly139→Arg, Asn182→Lys, Gln192→Arg and Asn193→Lys mutations separately into the HA of a clade-2 chicken isolate, A/chicken/Indonesia/N1/05 (CkInd). Either Gln192→Arg or Asn193→Lys, but not Gly139→Arg, enhanced the SA α 2,6Gal-binding affinity of CkInd to an extent similar to that observed for the HA of VN1194 (Fig. 3). Introduction of Asn182→Lys into the CkInd HA nearly abolished its binding to both SA α 2,3Gal and SA α 2,6Gal molecules, indicating the incompatibility of lysine at this position in the CkInd HA.

To address the structural basis for the acquisition of human receptor-binding capacity by the H5 HA, we determined the crystal structure of the VN1194 virus. The structure was solved by molecular replacement using coordinates from the A/duck/Singapore/95 HA (relevant crystallographic statistics are given in Supplementary Table 2). The backbone structure of the VN1194 HA is shown in Fig. 4a, superimposed on the backbone of the HA of A/duck/Singapore/95. The two structures are very similar (root mean-square deviation (r.m.s.d.) on all C α positions, 0.46 Å), reflecting their 94%

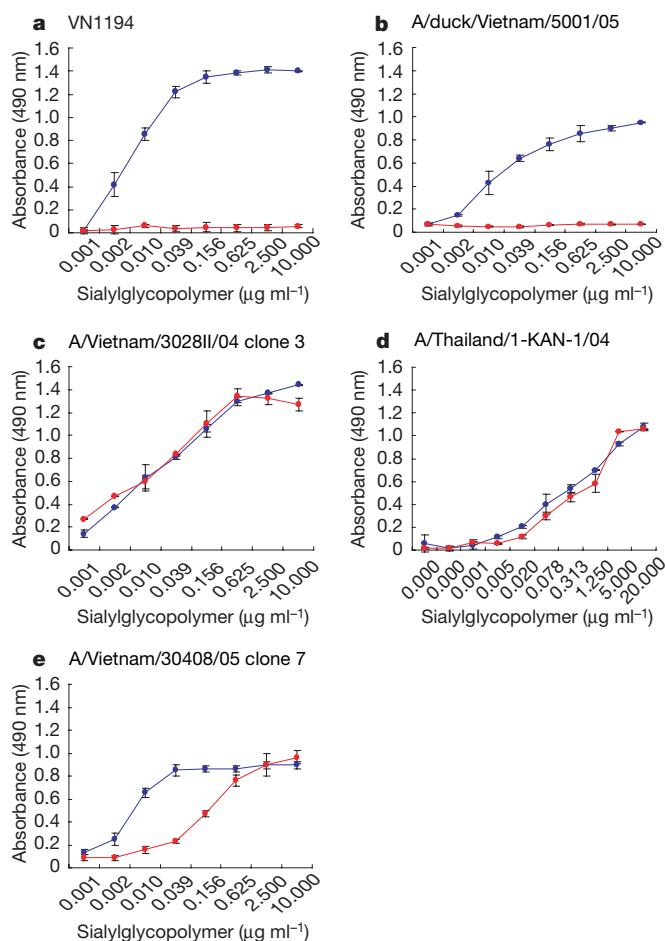


Figure 1 | Receptor-binding activity of H5N1 viruses. Direct binding of viruses to sialylglycopolymers containing either α 2,3-linked (blue) or α 2,6-linked (red) sialic acids was measured. **a**, Human isolate, Vietnam 2004. **b**, Virus isolated from duck. **c–e**, Human isolates with capacity to recognize both SA α 2,6Gal and SA α 2,3Gal. Data are the mean $\pm$ s.d. of triplicate experiments.

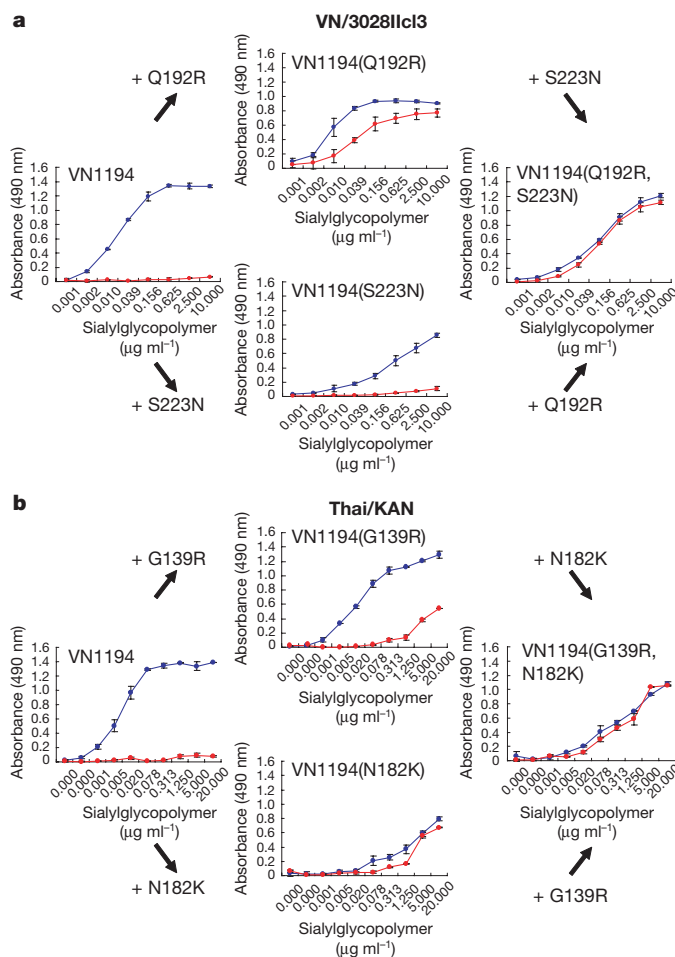


Figure 2 | Effect of HA mutations on the host cell receptor preference of the VN1194HA. Shown is the effect of HA mutations found in VN/3028IIc13 (**a**) and Thai/KAN (**b**) viruses. The mutations (in parentheses) were introduced first singly and then in combination. The HAs of the viruses possessing double mutations, VN1194(Q192R,S223N) in **a** and VN1194(G139R,N182K) in **b**, are identical to those of VN/3028IIc13 and Thai/KAN, respectively. Data are the mean $\pm$ s.d. of triplicate experiments.

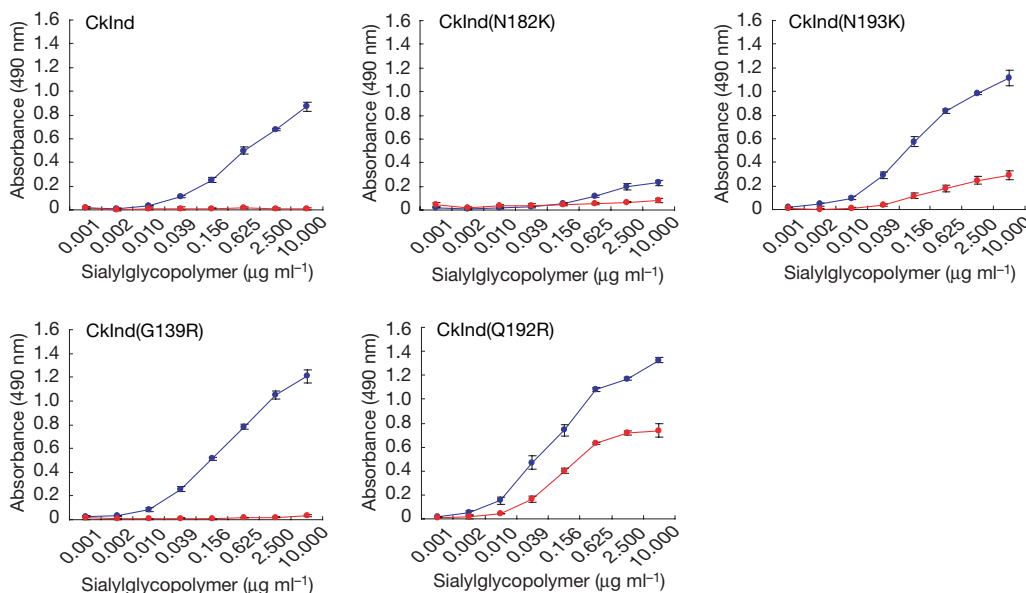


Figure 3 | Effect of mutations responsible for SA $\alpha 2,6$ Gal recognition by clade-1 HAs on a clade-2 HA. Mutations responsible for SA $\alpha 2,6$ Gal recognition by clade-1 HAs (in parentheses) were introduced into the HA of CkInd. The direct binding of each mutant virus to sialylglycopolymers containing either $\alpha 2,3$ -linked (blue) or $\alpha 2,6$ -linked (red) sialic acids was then measured at different concentrations of sialylglycopolymer. Data are the mean $\pm$ s.d. of triplicate experiments.

sequence identity. Figure 4b shows the receptor-binding domain, located on the globular head of the HA (Fig. 4a), of VN1194 superimposed on that of A/duck/Singapore/95 (ref. 9) and A/Vietnam/1203/04 (VN1203; ref. 10), which overlap with an r.m.s.d. of 0.46 Å and 0.5 Å, respectively, on the 175 C α positions of this domain. The overlap of the structure of the interhelical loop region of the HA2 of VN1194 and A/duck/Singapore/95 is shown in Fig. 4d. Again, the close correspondence of these structures presumably reflects their very high sequence identity. We note that the highly similar overall domain arrangement of the H5 HAs of VN1194 and A/duck/Singapore/95 (Fig. 4a) differs from the arrangement reported for

VN1203 (ref. 10). Further studies are needed to understand the basis of these differences.

Residues 75 in HA1 and 167 in HA2 are distant from the receptor-binding site, and the orientation of residues 139 and 193 precludes receptor contact (Fig. 4a). Clearly, further studies will be required to investigate the role of these residues. However, residues 182 and 192 are located at positions in the structure where it is feasible for them to make stabilizing interactions with sugars joined to sialic acid by $\alpha 2,6$ linkages (Fig. 4a, c). Mutations at residue 182 (the equivalent residue in the H3 HA is 186) have been linked to changes in receptor specificity^{9,11–13}, and mutating Asn 182 *in silico* to a lysine residue leads to

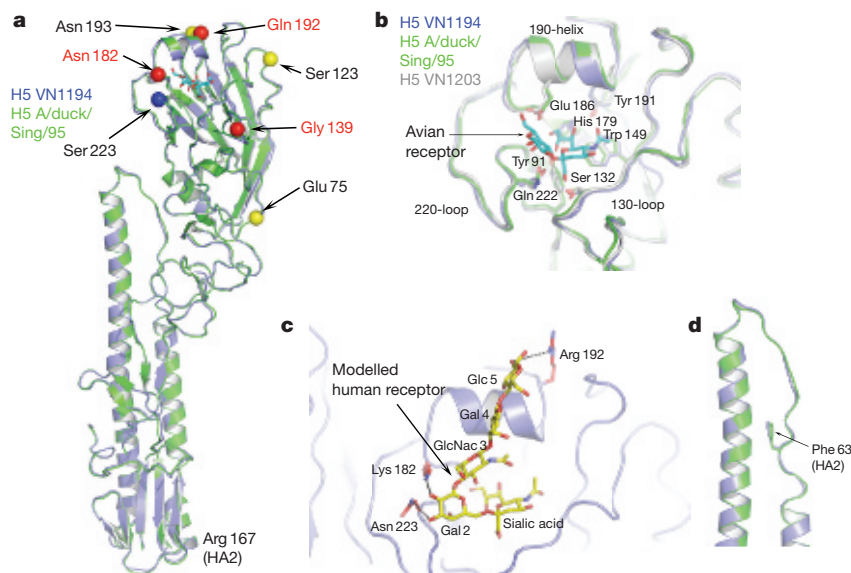


Figure 4 | Crystal structure of VN1194 H5 HA and the location of mutations conferring SA $\alpha 2,6$ Gal-binding capacity. **a**, Monomer from the crystal structure of the H5 HA from VN1194 (blue) determined at 2.8 Å, superimposed with a monomer of the H5 HA from A/duck/Singapore/95 (green). The mutations discussed in the text are indicated: residues in red represent positions where single substitutions affect receptor binding; those in yellow increase SA $\alpha 2,6$ Gal binding when introduced in combination with other changes. Asn 223 (blue) increases SA $\alpha 2,6$ Gal recognition in the presence of Arg 192. **b**, Superposition of the closely related receptor-binding domains of VN1194 (blue), A/duck/Singapore/95 (green) and VN1203 (grey) with the avian receptor analogue taken from the A/duck/Singapore/95

complex structure. The main secondary structure elements of the binding site (130-loop, 220-loop and 190-helix) and key binding site residues are indicated. **c**, More detailed view of the receptor-binding domain of the H5 HA of VN1194 with the human receptor analogue docked into the structure from its complex with the H1 HA. Three of the mutations discussed in the text are modelled in red: Asn182→Lys, Ser223→Asn and Gln192→Arg. Note that the residue numbering differs in the deposited coordinate file; for example, Asn 182 here is numbered 186 in the PDB file. **d**, Close up of the overlap between the interhelical regions of domain F of VN1194 (blue) and A/duck/Singapore/95 (green) and the orientation of the hydrophobic residue Phe 63 (HA2).

a potential hydrogen-bond interaction with the 2-OH of Gal-2 (Fig. 4c). Mutating Gln 192 (196 in the H3 HA) to an arginine residue with selection of a preferred rotamer (from the 'O' database)¹⁴ generates a conformation that places one of the guanidinium nitrogen atoms 4.5 Å from the 2-OH of Glc-5 (Fig. 4c); thus, the mutant HA may be capable of forming a hydrogen bond with human receptor moieties.

Serine 123 (128 in the H3 HA) is located on the turn leading into the 130-loop that forms the front edge of the binding site (Fig. 4b, c). Mutation of this residue could alter the orientation of the 130-loop and thus the attachment angle between the sialic acid and the next residue in the polysaccharide chain, thereby influencing the preference for the α 2,3 or α 2,6 linkage. Ser 223 is located close to the sialic-acid-binding site and *in silico* substitution of this residue with an asparagine leads to a conformation that places its side-chain nitrogen about 4 Å from the 3-OH of Gal-2 (Fig. 4c). This observation suggests that the mutant HA, with an asparagine at position 223, may be able to form a hydrogen bond with Gal-2 and thus influence SA α 2,6Gal recognition, although this amino acid substitution only slightly increases the SA α 2,6Gal-binding capacity of the VN1194 HA (Fig. 2a).

The influenza A viruses responsible for the pandemics of 1918, 1957 and 1968 all derived their HAs from avian viruses, which typically recognize SA α 2,3Gal (ref. 3). Yet, the HAs of early isolates from humans infected in these pandemics seem to have recognized SA α 2,6Gal in preference to SA α 2,3Gal (ref. 3), suggesting that conversion of the avian HA to one that can recognize SA α 2,6Gal-terminated polysaccharides on host cells is an important step in the generation of pandemic strains. This concept is supported by studies on the distribution of influenza virus receptors in the human airway^{15,16}. The critical amino acid substitutions involved in this shift of receptor recognition were residues 226 and 228 in the H2 and H3 HAs¹⁷ (equivalent to residues 222 and 224 in the H5 HA). The introduction of these mutations into the H5 HA permitted its binding to an α 2,6 glycan¹⁰, although neither change has been found in the HAs of H5N1 viruses isolated from humans.

In our study, both single and combined amino acid substitutions in the avian H5 HA mediated a shift to SA α 2,6Gal recognition. Moreover, two of these changes, lysine at position 182 and arginine at position 192, were present in the HAs of clade-2 H5N1 viruses isolated from two individuals in Azerbaijan and one individual in Iraq, but not in any of the more than 600 avian isolates examined. Although amino acid substitutions in viral proteins other than the HA, including PB2 (refs 18–20), may be needed to confer full pandemic status to an avian virus efficiently replicating in humans, the amino acid residues identified here may be selected during an early phase of human infection involving cells of the upper respiratory tract. Thus, such residues might provide useful molecular markers in assessments of H5N1 field isolates for their capacity to replicate in humans—an essential indicator of pandemic potential.

METHODS

Virus preparation. Viruses were grown in MDCK cells, which were maintained in minimal essential medium supplemented with 5% newborn calf serum (Sigma) and antibiotics at 37 °C in 5% CO₂. All experiments with live H5N1 virus were done in a biosafety level-3 containment laboratory.

Generation of viruses by reverse genetics. Reassortant viruses were generated with a plasmid-based reverse genetics system⁸. The viral complementary DNAs were cloned into a plasmid under control of the human polymerase I promoter and the mouse RNA polymerase I terminator (referred to as PolI plasmids). All viruses generated by reverse genetics possessed the neuraminidase of VN1194 and the internal genes of PR8. To generate the various HAs of human isolates registered in the Influenza Sequence Database (<https://www.flu.lanl.gov/>), we introduced mutations into pPolI-VN1194HA. Each construct was sequenced to verify the absence of unwanted mutations.

Receptor specificity assays. By using a solid-phase binding assay with the sodium salts of sialylglycopolymers (poly α -L-glutamic acid backbones containing N-acetylneuraminic acid linked to galactose through either an α -2,3 (Neu5Ac α 2,3Gal β 1,4GlcNAc β -pAP) or an α -2,6 (Neu5Ac α 2,6Gal β 1,

4GlcNAc β -pAP) bond) as described^{7,21,22}, we determined the direct binding capacity of the viruses for the sialylglycopolymers. In brief, polystyrene Universal-Bind microplates (Corning) were incubated with either of the two glycopolymers in PBS at 4 °C for 3 h, and then irradiated under ultraviolet light at 254 nm for 2 min. After removal of the glycopolymer solution, the plates were blocked with 0.1 ml of PBS containing 2% bovine serum albumin (Invitrogen) at room temperature for 1 h. After five washes with PBS, the plates were incubated in a solution containing influenza virus (128 haemagglutination units in PBS) at 4 °C for 12 h. After three washes with PBS, antibody to the virus was added to the plates. After 2 h of incubation at 4 °C, the plates were washed three times with ice-cold PBS and then incubated with horseradish peroxidase (HRP)-conjugated protein A (2000-fold dilution in PBS; Organon Teknica-Cappel) at 4 °C. After four washes with ice-cold PBS, the plates were incubated with O-phenylenediamine (Sigma) in PBS containing 0.01% H₂O₂ for 10 min at room temperature, and the reaction was stopped with 0.05 ml of 1 M HCl. Absorbance was determined at 490 nm.

To confirm the specificity of the assay, we carried out control experiments as follows. In brief, the glycopolymers were fixed on the plates as described above, 120 μ l of *A. ureafaciens* sialidase (80 mU ml⁻¹; Nacalai Tesque) was added and the plates were incubated at 37 °C for 16 h. After three washes with PBS, the plates were blocked with 0.1 ml of PBS containing 1% bovine serum albumin at 4 °C for 16 h. Removal of sialic acids from sialylglycopolymers was confirmed by incubating them with 50 μ l of biotinylated SNA specific for SA α 2,6Gal, biotinylated MALII specific for SA α 2,3Gal, or biotinylated MALI specific for both Gal β 1-4GlcNAc and SA α 2,3Gal β 1-4GlcNAc (all Vector Laboratories), at room temperature for 1 h. After three washes with PBS, the plates were incubated with HRP-conjugated streptavidin (Vector Laboratories) at room temperature for 1 h. After four washes with PBS, the plates were incubated with O-phenylenediamine in PBS containing 0.01% H₂O₂ for 10 min at room temperature, and the reaction was stopped with 0.05 ml of 1 M HCl. Absorbance was determined at 490 nm. Plates containing sialidase-treated glycopolymers and those containing poly α -L-glutamic acid backbones were used to confirm the lack of virus binding to the asialoglycopolymers or poly α -L-glutamic acid backbones.

Crystal structure determination. The H5N1 virus NIBRG-14, modified from A/Vietnam/1194/04 by removal of the polybasic cleavage site in HA, was obtained from the National Institute for Biological Standards and Control. HA was prepared by bromelain digestion of purified egg-grown virus for 90 min at 34 °C in 10 mM Tris-HCl (pH 8.0) and 5 mM mercaptoethanol (virus protein/bromelain ratio 10/1 w/w). The HA was purified as described²³. Crystallization conditions were screened by the sitting-drop vapour diffusion method using Crystal Clear strips (Douglas Instruments). The nanodrops were set up with 0.1 μ l of H5 protein solution (10 mg ml⁻¹) and 0.1 μ l of reservoir solution by using an Oryx1-6 robot (Douglas Instruments). Diffracting crystals were obtained from Index solution 57 (Hampton Research): namely, 30% (v/v) pentaerythritol ethoxylate, 50 mM ammonium sulphate and 50 mM Bis-Tris (pH 6.5). This solution also acted as a cryoprotectant. The H5 diffraction data were recorded on a Raxis4 detector (100- μ m scan) mounted on a Rigaku MicroMax 007 HF generator, integrated with Denzo and scaled with Scalepack²⁴. The structure was solved by molecular replacement using AmoRe²⁵ using the PDB file 1JSM (ref. 9) as the initial search model. Standard refinement was carried out with a combination of refmac5 (ref. 25) and CNS²⁶, together with manual model building with O¹². Molecular figures were created with Pymol (<http://pymol.sourceforge.net/>).

Received 20 August; accepted 21 September 2006.

- Webster, R. G., Peiris, M., Chen, H. & Guan, Y. H5N1 outbreaks and enzootic influenza. *Emerg. Infect. Dis.* **12**, 3–8 (2006).
- Enserink, M. Avian influenza. H5N1 moves into Africa, European Union, deepening global crisis. *Science* **311**, 932 (2006).
- Matrosovich, M. *et al.* Early alterations of the receptor-binding properties of H1, H2, and H3 avian influenza virus hemagglutinins after their introduction into mammals. *J. Virol.* **74**, 8502–8512 (2000).
- Rogers, G. N. & Paulson, J. C. Receptor determinants of human and animal influenza virus isolates: differences in receptor specificity of the H3 hemagglutinin based on species of origin. *Virology* **127**, 361–373 (1983).
- Rogers, G. N., Pritchett, T. J., Lane, J. L. & Paulson, J. C. Differential sensitivity of human, avian, and equine influenza A viruses to a glycoprotein inhibitor of infection: selection of receptor specific variants. *Virology* **131**, 394–408 (1983).
- Zambon, M. C. The pathogenesis of influenza in humans. *Rev. Med. Virol.* **11**, 227–241 (2001).
- Le, Q. M. *et al.* Avian flu: isolation of drug-resistant H5N1 virus. *Nature* **437**, 1108 (2005).
- Neumann, G. *et al.* Generation of influenza A viruses entirely from cloned cDNAs. *Proc. Natl Acad. Sci. USA* **96**, 9345–9350 (1999).

9. Ha, Y., Stevens, D. J., Skehel, J. J. & Wiley, D. C. X-ray structures of H5 avian and H9 swine influenza virus haemagglutinins bound to avian and human receptor analogs. *Proc. Natl Acad. Sci. USA* **98**, 11181–11186 (2001).
10. Stevens, J. *et al.* Structure and receptor specificity of the haemagglutinin from an H5N1 influenza virus. *Science* **312**, 404–410 (2006).
11. Eisen, M. B., Sabesan, S., Skehel, J. J. & Wiley, D. C. Binding of the influenza A virus to cell-surface receptors: structures of five haemagglutinin-sialyloligosaccharide complexes determined by X-ray crystallography. *Virology* **232**, 19–31 (1997).
12. Hardy, C. T. *et al.* Egg fluids and cells of the chorioallantoic membrane of embryonated chicken eggs can select different variants of influenza A (H3N2) viruses. *Virology* **211**, 302–306 (1995).
13. Gubareva, L. V. *et al.* Codominant mixtures of viruses in reference strains of influenza virus due to host cell variation. *Virology* **199**, 89–97 (1994).
14. Jones, T. A., Zhou, J. Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).
15. Shinya, K. *et al.* Avian flu: influenza virus receptors in the human airway. *Nature* **440**, 435–436 (2006).
16. Kuiken, T. *et al.* Host species barriers to influenza virus infections. *Science* **312**, 394–397 (2006).
17. Connor, R. J., Kawaoka, Y., Webster, R. G. & Paulson, J. C. Receptor specificity in human, avian, and equine H2 and H3 influenza virus isolates. *Virology* **205**, 17–23 (1994).
18. Clements, M. L. *et al.* Use of single-gene reassortant viruses to study the role of avian influenza A virus genes in attenuation of wild-type human influenza A virus for squirrel monkeys and adult human volunteers. *J. Clin. Microbiol.* **30**, 655–662 (1992).
19. Subbarao, E. K., London, W. & Murphy, B. R. A single amino acid in the PB2 gene of influenza A virus is a determinant of host range. *J. Virol.* **67**, 1761–1764 (1993).
20. Shinya, K. *et al.* PB2 amino acid at position 627 affects replicative efficiency, but not cell tropism, of Hong Kong H5N1 influenza A viruses in mice. *Virology* **320**, 258–266 (2004).
21. Shinya, K. *et al.* Characterization of a human H5N1 influenza A virus isolated in 2003. *J. Virol.* **79**, 9926–9932 (2005).
22. Totani, K. *et al.* Chemoenzymatic synthesis and application of glycopolymers containing multivalent sialyloligosaccharides with a poly(L-glutamic acid) backbone for inhibition of infection by influenza viruses. *Glycobiology* **13**, 315–326 (2003).
23. Ha, Y., Stevens, D. J., Skehel, J. J. & Wiley, D. C. H5 avian and H9 swine influenza virus haemagglutinin structures: possible origin of influenza subtypes. *EMBO J.* **21**, 865–875 (2002).
24. Otwinowski, Z. & Minor, W. in *Data Collection and Processing* (eds Sawyer, L., Isaacs, N. & Bailey, S.) 556–562 (SERC Daresbury Laboratory, Warrington, UK, 1993).
25. CCP4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
26. Brunger, A. T. *et al.* Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank K. Wells for technical assistance, and J. Gilbert for editing the manuscript. The NIMR contributors were responsible for the structural studies and for HA sequencing. This work was supported by CREST (Japan Science and Technology Agency); by grants-in-aid from the Ministry of Education, Culture, Sports, Science and Technology, Japan; by the Ministry of Health, Labour and Welfare, Japan; and by grants from the NIH, NIAID. Structural studies were supported by the UK MRC and by an International Partnership Research Award in Veterinary Epidemiology of the Wellcome Trust.

Author Contributions S.Y., Y.S., T.S., M.Q.L., C.A.N., Y.S.T., Y.M., T.H., T.S., M.K., T.U., T.M., Y.L., A.H. and Y.K. were responsible for the virological studies. L.F.H., D.J.S., R.J.R., S.J.G. and J.J.S. were responsible for the structural studies.

Author Information Coordinates for the H5 structure have been deposited in the Protein Data Bank under accession code 2IBX. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to Y.K. (kawaoka@ims.u-tokyo.ac.jp).

Visualization of transient encounter complexes in protein–protein association

Chun Tang¹, Junji Iwahara¹ & G. Marius Clore¹

Kinetic data on a number of protein–protein associations have provided evidence for the initial formation of a pre-equilibrium encounter complex that subsequently relaxes to the final stereo-specific complex¹. Site-directed mutagenesis^{2–4} and brownian dynamics simulations^{5–7} have suggested that the rate of association can be modulated by perturbations in charge distribution outside the direct interaction surfaces. Furthermore, rate enhancement through non-specific binding may occur by either a reduction in dimensionality⁸ or the presence of a short-range, non-specific attractive potential⁹. Here, using paramagnetic relaxation enhancement, we directly demonstrate the existence and visualize the distribution of an ensemble of transient, non-specific encounter complexes under equilibrium conditions for a relatively weak protein–protein complex between the amino-terminal domain of enzyme I and the phosphocarrier protein HPr. Neither the stereo-specific complex¹⁰ alone nor any single alternative conformation can account fully for the intermolecular paramagnetic relaxation enhancement data. Restrained rigid-body simulated annealing refinement against the paramagnetic relaxation enhancement data enables us to obtain an atomic probability distribution map of the non-specific encounter complex ensemble that qualitatively correlates with the electrostatic surface potentials on the interacting proteins. Qualitatively similar results are presented for two other protein–protein complexes.

The association of the N-terminal domain of enzyme I (EIN) and the phosphocarrier protein HPr (dissociation constant, $K_d \approx 10 \mu\text{M}$), the first binary complex in the bacterial phosphotransferase system¹¹, is in fast exchange on the chemical shift scale¹⁰. The structure of the stereospecific EIN–HPr complex has been solved by NMR on the basis of nuclear Overhauser enhancement (NOE) and residual dipolar coupling (RDC) data¹⁰. These data are fully consistent with a single, unique conformation that readily accounts for the phosphoryl transfer reaction between the two proteins¹⁰. However, neither the NOE nor the RDCs are sensitive to the presence of low-population ($\leq 10\%$) intermediates. To detect such intermediates, we introduced a paramagnetic label at three sites on HPr, one at a time, and measured the transverse paramagnetic relaxation enhancement (PRE) rates, T_2 , of the backbone amide protons ($^1\text{H}_\text{N}$) of EIN. In a fast exchanging system, the observed value of T_2 is the weighted average of the T_2 values for the various states present in solution^{12,13}. Because T_2 is dependent on the sixth root of the distance ($\propto r^{-6}$) between the unpaired electron on the paramagnetic centre and the observed proton, and because the T_2 rates at short distances are very large owing to the large magnetic moment of the unpaired electrons, low-population intermediates can be detected. Glu 5, Glu 25 and Glu 32 of HPr, which are all located outside the specific interaction surface with EIN, were substituted individually by a cysteine residue, which was then conjugated to EDTA through a disulphide linkage to yield a (cysteaminyl-EDTA)-Cys adduct. The latter is chelated to

either Mn^{2+} (paramagnetic state) or Ca^{2+} (diamagnetic state)¹⁴. These modifications do not change the net charge of HPr, nor do they perturb the binding equilibrium with EIN.

The intramolecular $^1\text{H}_\text{N}$ - T_2 rates for HPr in the EIN–HPr complex are fully consistent with the static structure of HPr, with an overall PRE Q-factor¹⁵ (for all three paramagnetic sites combined) of 0.18 and a correlation coefficient of 0.94 (Supplementary Fig. S1). (The Q-factor is a quantitative measure of agreement between observed and calculated T_2 rates and is given by equation (2) in the Methods section.) A comparison, however, of the observed intermolecular $^1\text{H}_\text{N}$ - T_2 rates measured on EIN with those back-calculated from the structure of the stereospecific complex as a function of residue number reveals regions with large discrepancies (Fig. 1). The correlation between observed and calculated intermolecular $^1\text{H}_\text{N}$ - T_2 rates is very poor, with an overall Q-factor of 0.61 (Fig. 2a). In the stereospecific complex, the intermolecular contacts predominantly involve helices $\alpha 2$ and $\alpha 2'$, the carboxy-terminal end of helix $\alpha 3$ and the N-terminal end of helix $\alpha 4$ of the α -domain of EIN, and helices $\alpha 1$ and $\alpha 2$ of HPr¹⁰. For each paramagnetic site, there are regions with large observed intermolecular $^1\text{H}_\text{N}$ - T_2 rates that are well predicted by the stereospecific complex (that is, they are in close proximity to the paramagnetic labels): namely, residues 110–137, 50–92 and 73–140 for Glu5→Cys, Glu25→Cys and Glu32→Cys, respectively, within the α -domain of EIN (Fig. 1). However, there are many other residues within the α -domain of EIN that are far away ($\geq 25 \text{ \AA}$) from the paramagnetic labels but show large $^1\text{H}_\text{N}$ - T_2 rates that are inconsistent with the structure of the specific complex: namely, residues 59–97, 105–124 and 20–71 for Glu5→Cys, Glu25→Cys and Glu32→Cys, respectively (Fig. 1). In addition, there are regions in the α/β domain, further away from the paramagnetic centres, where the agreement is poor to moderate: namely, residues 23–37, 183–189 and 241–249 for Glu25→Cys, and residues 184–189 and 232–249 for Glu32→Cys (Fig. 1). The discrepancies cannot be explained by a solvent PRE effect^{16,17} involving diffusion and random collisions between EIN and the paramagnetically labelled HPr because, at the relatively low concentrations used ($\sim 300 \mu\text{M}$), no significant PRE effects ($> 2 \text{ s}^{-1}$) were observed for a control protein (that does not interact with HPr) on addition of paramagnetically labelled HPr. Thus, the observed intermolecular PRE data provide unambiguous qualitative evidence for the existence of lowly populated ($\leq 10\%$) minor species in rapid exchange with the final stereospecific complex (see Supplementary Information).

A semi-quantitative depiction of the minor species was obtained by using restrained rigid-body simulated annealing refinement¹⁸ to minimize the difference between observed and calculated $^1\text{H}_\text{N}$ - T_2 rates for all three paramagnetic sites simultaneously by representing the minor non-specific states by an ensemble of HPr molecules (with atomic overlap between HPr molecules allowed because the ensemble reflects a population distribution). We carried out 100 independent

¹Laboratory of Chemical Physics, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland 20892-0520, USA.

calculations with ensemble sizes N ranging from 1 to 20, varying the percentage of the minor species (p_{minor}) relative to the stereospecific complex. Complete cross-validation¹⁹, leaving out random portions of 10% of the complete PRE data set from all three sites, was done to assess how well the test PRE data sets (10% excluded from the refinement) are predicted by the working data sets (90% included in the refinement). Introduction of a single minor species results in only a modest decrease in the Q -factor, but, as the number of conformers is increased, a large decrease in the Q -factor is observed (Fig. 2c). Complete cross-validation indicates that the optimal number of conformers required to satisfy the data is 10–20 (Fig. 2c) and that the improvement in Q -factor is not a result of over-fitting the data. Thus, the minor species comprises many non-specific binding states that reflect an ensemble of transient encounter complexes.

The Q -factor decreases rapidly as p_{minor} is increased to 10% (Fig. 2d), a value that is still consistent with the NOE and RDC data, and thereafter slowly levels off. The Q -factor obtained by averaging the $^1\text{H}_\text{N}$ - Γ_2 rates over all ensembles (that is, the ensemble of ensembles average, Q_{ee}) is systematically smaller than the average Q -factor for the individual ensembles (Q_e). This is due to the stochastic rather than unique configuration of states within each ensemble, such that averaging over all ensembles affords a better representation of the data (Fig. 2c, d). For $N = 20$ and $p_{\text{minor}} = 10\%$, Q_{ee} has a value of 0.21 and the correlation coefficient between observed and calculated $^1\text{H}_\text{N}$ - Γ_2 rates is 0.97 (Fig. 2b), which is comparable to the agreement observed for the intramolecular PRE data (Supplementary Fig. S1). The overall population of minor non-specific encounter complexes may seem to be relatively high but is perhaps not unexpected for a relatively weak protein–protein association. However, the occupancy of any individual conformer in the ensemble of non-specific

encounter complexes is very small and has therefore eluded structural characterization by any other experimental method.

The minor species were visualized by using a reweighted atomic probability density map²⁰ derived from 100 independent calculations using an ensemble size of $N = 20$ (Fig. 3a–c, left). The distribution of non-specifically bound HPr molecules on the surface of EIN is essentially continuous but non-uniform. The probability density in the region of the specific complex is low, indicating that alternative modes of binding confined to the contact surfaces involved in the stereospecific complex do not significantly contribute to the discrepancy between the observed PREs and those calculated on the basis of the specific complex (Supplementary Figs S2 and S3).

Outside the specific interaction surface, the distribution of HPr molecules (Fig. 3a–c, left) is qualitatively correlated with the negative electrostatic potential isosurface²¹ of EIN (Fig. 3a–c, right). EIN is an acidic protein and large areas of its surface are swathed by a negative electrostatic potential. About half the surface of HPr, including the interaction surface used in the specific complex, is positively charged, and the remaining surface (formed predominantly by the β -sheet) is negatively charged. The HPr atomic probability density map heavily populates regions of EIN with highly negative electrostatic potentials, to a lesser extent around the C terminus of EIN, and minimally on the back of EIN (Supplementary Fig. S4a). Note that the HPr conformers located at the α/β domain are, on average, ~ 50 Å from HPr in the stereospecific complex. In all cases, HPr preferentially uses its positively charged surface to interact with the negatively charged regions of EIN (Supplementary Fig. S5c). Thus, the modifications used to introduce paramagnetic groups on HPr do not affect the formation of non-specific encounter complexes. These findings suggest that the formation of non-specific EIN–HPr encounter complexes is

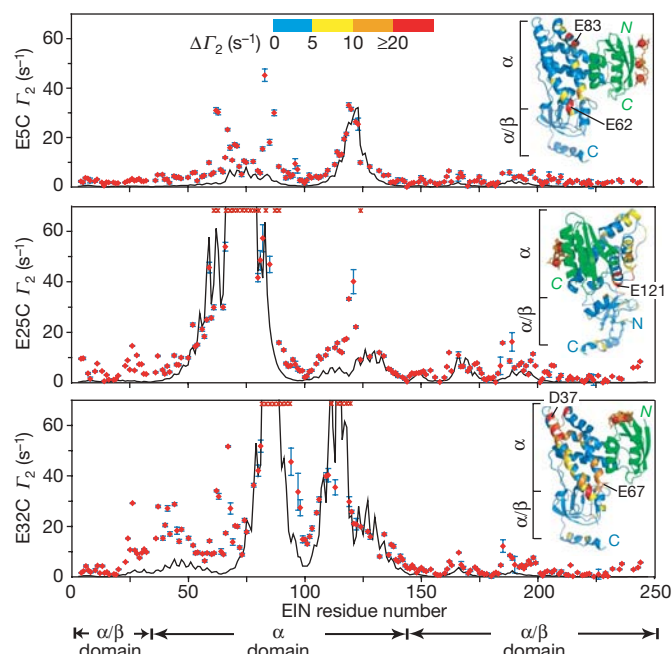


Figure 1 | Observed and calculated intermolecular PREs for the EIN–HPr complex. Shown is a comparison of observed intermolecular $^1\text{H}_\text{N}$ - Γ_2 rates (red diamonds) with those back-calculated (black lines) from the stereospecific complex; paramagnetic labels (EDTA- Mn^{2+}) were introduced one at a time at three sites (E5C, E25C and E32C) on HPr. Red crosses indicate residues with $^1\text{H}_\text{N}/^{15}\text{N}$ cross-peaks that are broadened beyond detection by PRE. Insets show the structure of the stereospecific EIN–HPr complex with EIN colour coded according to the difference, $\Delta\Gamma_2$, between the observed and calculated intermolecular $^1\text{H}_\text{N}$ - Γ_2 rates for each paramagnetic site. (HPr, green; three-member ensemble representation of EDTA- Mn^{2+} with EDTA and linkage, orange, and Mn^{2+} , red spheres).

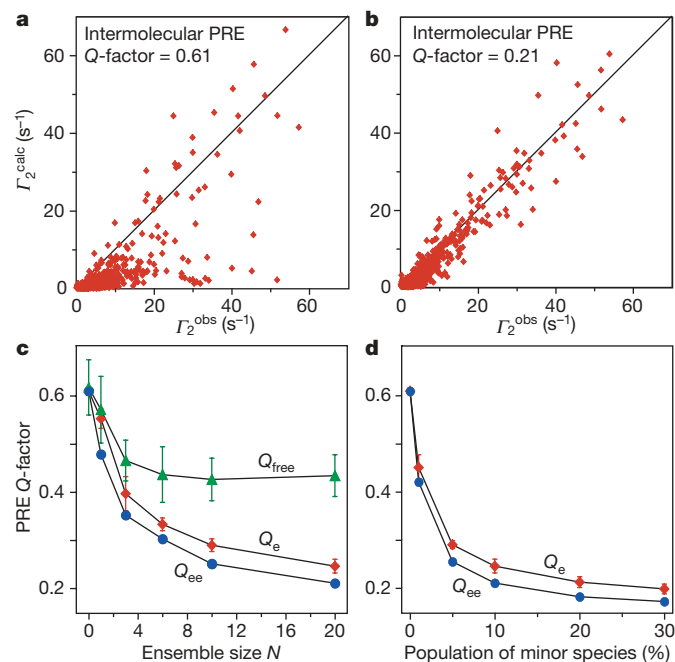


Figure 2 | Ensemble refinement and intermolecular PRE Q -factor.

a, b, Correlation between observed intermolecular $^1\text{H}_\text{N}$ - Γ_2 rates (501 data points) and those calculated from the structure of the stereospecific complex either alone (**a**) or with the addition of an ensemble of $N = 20$ to represent the non-specific encounter complex (**b**; $p_{\text{minor}} = 10\%$; averaged over 100 independent calculations). **c,** Dependence of the working (Q_e , red; Q_{ee} , blue) and cross-validated (Q_{free} , green) Q -factors on ensemble size N ($p_{\text{minor}} = 10\%$). **d,** Dependence of Q_e and Q_{ee} on p_{minor} ($N = 20$). The points in **c** and **d** at $N = 0$ and $p_{\text{minor}} = 0$, respectively, represent control calculations in which $p_{\text{minor}} = 0$ and the position of HPr for the single specific complex is optimized by rigid-body simulated annealing to satisfy the intermolecular PRE data. Error bars indicate the s.d.

predominantly driven by weak non-specific electrostatic attractions between the two molecules.

The total solvent accessible surface area (ASA) buried at the interfaces of the non-specific encounter complexes is, on average, an order of magnitude smaller than that of the stereospecific complex (1,945 Å²; Fig. 3d). The non-specific interfaces are also much less compact, with gap indices (defined as the ratio of gap volume to buried interface ASA²²) many times larger than that of the stereospecific complex (2.1 Å; Fig. 3d). In addition, the non-specific interfaces are more planar than the stereospecific one, indicative of the absence of lock-and-key binding (Supplementary Table S1). These observations are consistent with the correlation between the spatial distribution of non-specific encounter complexes and electrostatic potential isosurface, because electrostatic interactions are relatively long range (with a $1/r$ distance dependence) and do not necessarily require van der Waals contact. The role of electrostatic interactions is reinforced by the observation that the interfacial composition of charged residues is increased, whereas that of uncharged polar and

non-polar residues is decreased in the non-specific encounter complexes relative to the stereospecific complex (Supplementary Fig. S4b).

Once a non-specific encounter complex is formed by weak non-specific electrostatic interactions, HPr can carry out a two-dimensional search on the surface of EIN, eventually falling into a narrow energy funnel that leads directly to the stereospecific complex characterized by an array of complementary van der Waals and electrostatic interactions. The observation that the region on EIN comprising the specific interaction surface for HPr is only minimally occupied by non-specific encounter complexes (Fig. 3a) indicates that once HPr reaches this region formation of the stereospecific complex occurs with high probability.

The direct detection of non-specific encounter complexes by PRE is not confined to the EIN–HPr complex. We observed similar effects for two other weak ($K_d \approx 30$ – $50 \mu\text{M}$), fast-exchanging protein–protein complexes of the bacterial phosphotransferase system, IIA^{Mannitol}–HPr (Fig. 4) and IIA^{Mannose}–HPr (Supplementary Fig. S6), the stereospecific structures of which have been solved^{23,24}. With the paramagnetic label on HPr located at Glu5→Cys, on the opposite face to the stereospecific binding site, there are regions in both complexes where the observed intermolecular $^1\text{H}_N$ - T_2 rates are much larger than those back-calculated from the structures of the stereospecific complexes. The largest discrepancies involve acidic residues of IIA^{Mannitol} and IIA^{Mannose}, the distribution of which correlates qualitatively with the negative electrostatic potential on the surface of these proteins. Thus, in all likelihood the observations

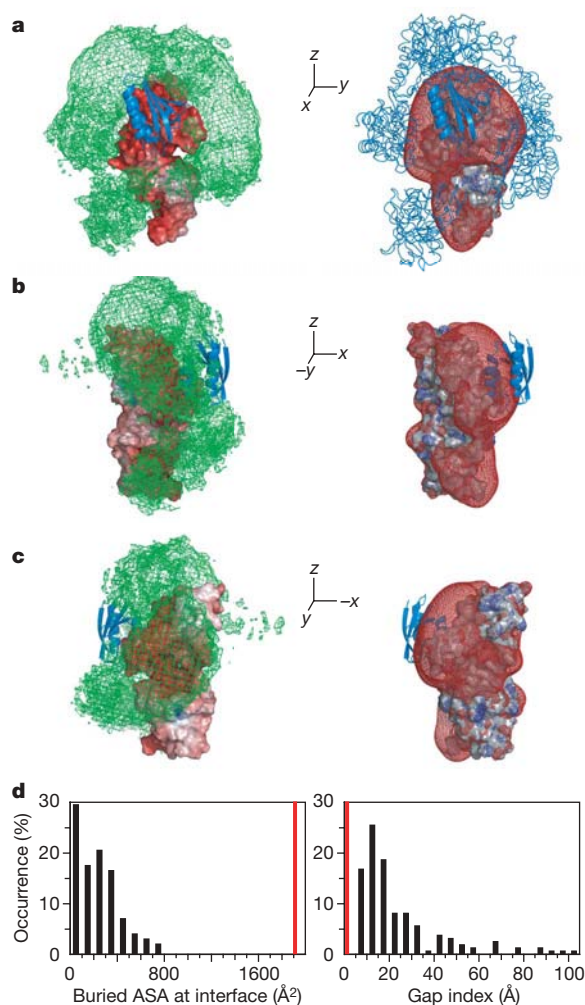


Figure 3 | Characterization of non-specific EIN–HPr encounter complexes. **a–c**, Left, overall distribution of HPr molecules obtained from 100 calculations ($N = 20$) displayed as a reweighted atomic probability density map²⁰ (plotted at a threshold of 20% maximum, green) on the molecular surface of EIN (colour coded by electrostatic potential, ± 8 kT). Right, molecular surface of EIN in grey with the electrostatic potential isosurface of EIN, calculated at ± 5 kT, displayed as red (negative) and blue (positive) meshes. The disposition of HPr molecules (blue tubes) in a typical ensemble of $N = 20$ is shown in the right panel of **a**. The location of HPr in the stereospecific complex is shown as a blue ribbon in all panels. **d**, Histograms of interface-buried ASA and gap index for the non-specific encounter complexes. Red lines indicate values for the stereospecific complex.

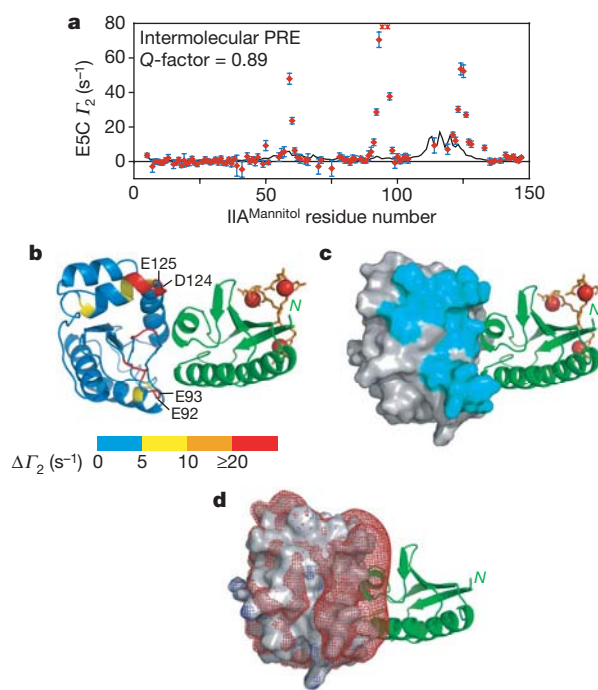


Figure 4 | Observed and calculated PRE $^1\text{H}_N$ - T_2 values for the IIA^{Mannitol}–HPr complex. EDTA- Mn^{2+} paramagnetic labels were introduced onto HPr at E5C. **a**, Comparison of observed intermolecular $^1\text{H}_N$ - T_2 rates (red diamonds) with those back-calculated (black lines) from the stereospecific complex. Red crosses indicate residues with $^1\text{H}_N/^{15}\text{N}$ cross-peaks that are broadened beyond detection by PRE; error bars indicate the s.d. **b**, Structure of the stereospecific complex with IIA^{Mannitol} colour coded according to the difference, ΔT_2 , between observed and calculated intermolecular $^1\text{H}_N$ - T_2 rates. Colouring for HPr and EDTA- Mn^{2+} as in Fig. 1 insets. **c**, Molecular surface representation of IIA^{Mannitol} in the stereospecific complex. Residues of IIA^{Mannitol} that show large ΔT_2 are coloured cyan. **d**, Electrostatic potential isosurface of IIA^{Mannitol} (± 5 kT in blue and red, respectively). In **c** and **d**, HPr is shown as a green ribbon. The same view is shown in **b–d**.

reported here reflect a general phenomenon of protein–protein association in which the initial formation of non-specific encounter complexes through long-range electrostatic interactions (possibly supplemented by short-range van der Waals interactions) facilitates the rapid formation of a stereospecific complex by reducing the dimensionality of the search process.

METHODS

Sample preparation and NMR spectroscopy. Details of sample preparation are provided in Supplementary Information. We acquired PRE data on a Bruker DRX-600 spectrometer as described¹⁵ (see Supplementary Information).

Back-calculation of PREs. Intra- and intermolecular PREs were back-calculated from the structures of the stereospecific complexes by using a three-conformer ensemble representation for the EDTA-Mn²⁺ groups to account for their flexibility (see Supplementary Information)¹⁵.

Ensemble refinement against intermolecular PREs. Refinement against the intermolecular PREs for the EIN–HPr complex was carried out by rigid body refinement with Xplor-NIH¹⁸ subject to a target function comprising the PRE data for all three paramagnetic sites¹⁵, a quartic van der Waals repulsion term²⁵ (to prevent atomic overlap between EIN and HPr) and a very weak radius of gyration restraint²⁶ (to ensure that each member of the ensemble makes at least some intermolecular contacts). Calculations were carried out either by keeping EIN fixed and allowing an ensemble of HPr molecules to rotate and translate, or by the converse (HPr fixed and an ensemble of EIN molecules), with essentially identical results (Supplementary Fig. S5). Further details of the calculations are provided in Supplementary Information. The calculated PRE for residue *i*, $\Gamma_2^{\text{calc}}(i)$, is given by

$$\Gamma_2^{\text{calc}}(i) = \lambda \Gamma_2^{\text{specific}}(i) + (1 - \lambda) \sum_{j=1}^N \Gamma_2^{\text{non-specific}}(i, j) / N \quad (1)$$

where λ is the fraction of the stereospecific complex ($\lambda = 1 - p_{\text{minor}}$), *N* is the ensemble size of the non-specific encounter complex, $\Gamma_2^{\text{specific}}(i)$ is the calculated Γ_2 rate for residue *i* in the stereospecific complex, and $\Gamma_2^{\text{non-specific}}(i, j)$ is the calculated Γ_2 rate for residue *i* of member *j* of the non-specific encounter complex ensemble. The PRE *Q*-factor is a measure of the agreement between observed and calculated values of Γ_2 and is given by:

$$Q = \left[\sum_i \{ \Gamma_2^{\text{obs}}(i) - \Gamma_2^{\text{calc}}(i) \}^2 / \sum_i \Gamma_2^{\text{obs}}(i)^2 \right]^{1/2} \quad (2)$$

where $\Gamma_2^{\text{obs}}(i)$ is the observed Γ_2 rate for residue *i*. Two *Q*-factors are reported: Q_e is the average *Q*-factor $\langle Q \rangle$ for all calculated *n* ensembles, with $\Gamma_2^{\text{calc}}(i)$ computed by equation (1); Q_{ee} is the ensemble of ensembles average *Q*-factor, computed by using the average value of $\Gamma_2^{\text{calc}}(i)$ over all *n* ensembles, with $\Gamma_2^{\text{calc}}(i)$ given by:

$$\Gamma_2^{\text{calc}}(i) = \lambda \Gamma_2^{\text{specific}}(i) + (1 - \lambda) \sum_{k=1}^n \sum_{j=1}^N \Gamma_2^{\text{non-specific}}(i, j, k) / nN \quad (3)$$

Analysis of complexes. Electrostatic potentials were calculated with APBS²¹ and are displayed in PyMol²⁷. Solvent ASA buried at the interface and gap volume were calculated with Xplor-NIH¹⁸ and SURFNET²⁸, respectively.

Received 23 May; accepted 4 September 2006.

Published online 15 October 2006.

1. Fersht, A. R. *Structure and Mechanism in Protein Science: a Guide to Enzyme Catalysis and Protein Folding* (Freeman & Co, New York, 1999).
2. Schreiber, G. & Fersht, A. R. Rapid electrostatically assisted association of proteins. *Nature Struct. Biol.* **3**, 427–431 (1996).
3. Vijaykumar, M. et al. Electrostatic enhancement of diffusion-controlled protein–protein association: comparison of theory and experiment on barnase and barstar. *J. Mol. Biol.* **278**, 1015–1024 (1998).
4. Selzer, T., Albeck, S. & Schreiber, G. Rational design of faster associating and tighter binding protein complexes. *Nature Struct. Biol.* **7**, 537–541 (2000).
5. Northrup, S. H., Boles, J. O. & Reynolds, J. C. L. Brownian dynamics of cytochrome c and cytochrome c peroxidase association. *Science* **241**, 67–70 (1988).
6. Gabdoulline, R. R. & Wade, R. C. Biomolecular diffusional association. *Curr. Opin. Struct. Biol.* **12**, 204–213 (2002).

7. Spaar, A., Dammer, C., Gabdoulline, R. R., Wade, R. C. & Helms, V. Diffusional encounter of barnase and barstar. *Biophys. J.* **90**, 1913–1924 (2006).
8. Adams, G. & Debruck, M. in *Structural Chemistry and Molecular Biology* (eds Rich, A. & Davidson, N.) 198–215 (Freeman & Co, San Francisco, CA, 1968).
9. Zhou, H.-X. & Szabo, A. Enhancement of association rates by nonspecific binding to DNA and cell membranes. *Phys. Rev. Lett.* **93**, 178101 (2004).
10. Garrett, D. S., Seok, Y.-J., Peterkofsky, A., Gronenborn, A. M. & Clore, G. M. Solution structure of the 40,000 *M_r* phosphoryl transfer complex between the N-terminal domain of enzyme I and HPr. *Nature Struct. Biol.* **6**, 166–173 (1999).
11. Postma, P. W., Lengeler, J. W. & Jacobson, G. R. in *Escherichia coli and Salmonella: Cellular and Molecular Biology* (ed. Neidhardt, F. C.) 1149–1174 (ASM, Washington DC, 1996).
12. Iwahara, J. & Clore, G. M. Detecting transient intermediates in macromolecular binding by paramagnetic NMR. *Nature* **440**, 1227–1230 (2006).
13. Iwahara, J., Schwieters, C. D. & Clore, G. M. Characterization of nonspecific protein–DNA interactions by ¹H paramagnetic relaxation enhancement. *J. Am. Chem. Soc.* **126**, 12800–12808 (2004).
14. Ebright, Y. W., Chen, Y., Pendergrast, P. S. & Ebright, R. H. Incorporation of an EDTA–metal complex at a rationally selected site within a protein: application to EDTA–iron DNA affinity cleaving with catabolite gene activator protein (CAP) and Cro. *Biochemistry* **31**, 10664–10670 (1992).
15. Iwahara, J., Schwieters, C. D. & Clore, G. M. Ensemble approach for NMR structure refinement against ¹H paramagnetic relaxation enhancement data arising from a flexible paramagnetic group attached to a macromolecules. *J. Am. Chem. Soc.* **126**, 5879–5896 (2004).
16. Donaldson, L. W. et al. Structural characterization of proteins with an attached ATCUN motif by paramagnetic relaxation enhancement NMR spectroscopy. *J. Am. Chem. Soc.* **123**, 9843–9847 (2001).
17. Pintacuda, G. & Otting, G. Identification of protein surfaces by NMR measurements with a paramagnetic Gd(III) chelate. *J. Am. Chem. Soc.* **124**, 372–373 (2002).
18. Schwieters, C. D., Kuszewski, J. J. & Clore, G. M. Using Xplor-NIH for NMR molecular structure determination. *Prog. NMR Spectrosc.* **48**, 47–62 (2006).
19. Brünger, A. T., Clore, G. M., Gronenborn, A. M. & Nilges, M. Assessing the quality of solution nuclear magnetic resonance structures by complete cross-validation. *Science* **261**, 328–331 (1993).
20. Schwieters, C. D. & Clore, G. M. Reweighted atomic densities to represent ensembles of NMR structures. *J. Biomol. NMR* **23**, 221–225 (2002).
21. Baker, N. A., Sept, D., Joseph, S., Holst, M. J. & McCammon, J. A. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl Acad. Sci. USA* **98**, 10037–10041 (2001).
22. Jones, S. & Thornton, J. M. principles of protein–protein interactions. *Proc. Natl Acad. Sci. USA* **93**, 13–20 (1996).
23. Cornilescu, G. et al. Solution structure of the phosphoryl transfer complex between the cytoplasmic A domain of the mannitol transporter II^{Mannitol} and HPr of the *Escherichia coli* phosphotransferase system. *J. Biol. Chem.* **277**, 42289–42298 (2002).
24. Williams, D. C., Cai, M., Suh, Y.-J., Peterkofsky, A. & Clore, G. M. Solution NMR structure of the 48 kDa IIA^{Mannose}–HPr complex of the *Escherichia coli* mannose phosphotransferase system. *J. Biol. Chem.* **280**, 20775–20784 (2005).
25. Nilges, M., Gronenborn, A. M., Brünger, A. T. & Clore, G. M. Determination of three-dimensional structures of proteins by simulated annealing with interproton distance restraints: application to crambin, potato carboxypeptidase inhibitor and barley serine proteinase inhibitor 2. *Protein Eng.* **2**, 27–38 (1988).
26. Kuszewski, J., Gronenborn, A. M. & Clore, G. M. Improving the packing and accuracy of NMR structures with a pseudopotential for the radius of gyration. *J. Am. Chem. Soc.* **121**, 2337–2338 (1999).
27. DeLano, W. L. The PyMOL Molecular Graphics System (DeLano Scientific, San Carlos, CA, USA, 2002).
28. Laskowski, R. A. SURFNET: a program for visualizing molecular surfaces, cavities and intermolecular interactions. *J. Mol. Graph.* **13**, 323–330 (1995).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank C. Schwieters, A. Szabo and C. Bewley for discussions; and C. Byeon for assistance with initial sample preparation. This work was supported by funds from the Intramural Program of the NIH, NIDDK and the AIDS Targeted Antiviral program of the Office of the Director of the NIH (to G.M.C.).

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to G.M.C. (marisc@intra.niddk.nih.gov).

Bacterial chromatin organization by H-NS protein unravelled using dual DNA manipulation

Remus T. Dame^{1*}, Maarten C. Noom^{1*} & Gijs J. L. Wuite¹

Both prokaryotic and eukaryotic organisms contain DNA bridging proteins, which can have regulatory or architectural functions¹. The molecular and mechanical details of such proteins are hard to obtain, in particular if they involve non-specific interactions. The bacterial nucleoid consists of hundreds of DNA loops, shaped in part by non-specific DNA bridging proteins such as histone-like nucleoid structuring protein (H-NS), leucine-responsive regulatory protein (Lrp) and SMC (structural maintenance of chromosomes) proteins^{2,3}. We have developed an optical tweezers instrument that can independently handle two DNA molecules, which allows the systematic investigation of protein-mediated DNA–DNA interactions. Here we use this technique to investigate the abundant non-specific nucleoid-associated protein H-NS, and show that H-NS is dynamically organized between two DNA molecules in register with their helical pitch. Our optical tweezers also allow us to carry out dynamic force spectroscopy on non-specific DNA binding proteins and thereby to determine an energy landscape for the H-NS–DNA interaction. Our results explain how the bacterial nucleoid can be effectively compacted and organized, but be dynamic in nature and accessible to DNA-tracking motor enzymes. Finally, our experimental approach is widely applicable to other DNA bridging proteins, as well as to complex DNA interactions involving multiple DNA molecules.

The chromosomal DNA of bacteria is not confined to an envelope-enclosed organelle like the eukaryotic nucleus. Instead, it is folded into a compact nucleoid body. The folding is attributed in part to interactions with abundant nucleoid-associated proteins (NAPs)^{2,3}. The binding of NAPs to DNA is apparently compatible with DNA-tracking motor proteins such as RNA polymerase (RNAP). Moreover, the nucleoid body has to be constantly remodelled in response to environmental signals. How effective compaction by NAPs is compatible with these constraints is not well understood. H-NS is believed to have an essential role in the organization and compaction of bacterial chromatin^{2,4,5} and acts by bridging DNA duplexes^{6,7}. Large regions of DNA can be bridged by adjacent H-NS dimers or multimers in a presumably cooperative manner^{2,6–8}. Owing to its chromosome-wide structural impact, H-NS also acts as a global regulator of gene expression, and is essential for environmental adaptation and bacterial virulence^{4,5}. There is limited knowledge about the structural and kinetic aspects of these important H-NS–DNA interactions.

Details of protein-mediated DNA duplex bridging in general (but in particular by non-specific proteins) are hard to obtain with current bulk techniques. Also, using a single DNA molecule to probe such DNA binding proteins^{9–11} can be problematic, as the molecular details of DNA bridging are obscured by the large-scale effects of multiple-site DNA binding (compaction; see Supplementary Fig. 1). We have designed and built an optical tweezers instrument with

which we can simultaneously trap four micron-sized polystyrene beads (see Methods) that are attached to the ends of two single DNA molecules (Fig. 1a). Using two DNA molecules in an extended configuration allows us to address fundamental properties of DNA bridging proteins and their interaction with DNA. Finally, it opens up the possibility of high-throughput dynamic force spectroscopy on non-specific DNA binding proteins, yielding an energy landscape for their interaction with DNA.

With this instrument we have investigated the stability of the bridges in a DNA–H-NS–DNA assembly (H-NS–DNA₂) by exerting force through the duplexes. By building up a shearing force (Fig. 1b, inset) we show that an H-NS–DNA₂ complex is formed by dimeric or multimeric H-NS binding. When the beads are moved apart at a constant speed of 22 nm s^{−1}, the force rises to several tens of pN and then drops to zero after complete separation of the bridged region (Fig. 1b). If the bead movement is stopped after reaching a pre-set force, the force gradually decreases, presumably owing to the disruption of bridges at the extremities of the bridged regions (Fig. 1b). If we allow H-NS to coat both DNA duplexes before we establish a crossed configuration, we do not observe H-NS-mediated DNA–DNA interactions (Fig. 1b). This finding indicates that protein–protein interactions are not responsible for bridge formation, but that instead each H-NS dimer or multimer interacts with two duplexes simultaneously. The application of an unzipping force (Fig. 1c, inset) gives rise to a rugged force landscape (Fig. 1c). The sign of the slope in such graphs is defined by the relative rates of bead movement and H-NS unbinding and, as such, reflects the local density of H-NS bridges. Depending on the bead's speed, a different force regime is associated with disruption: up to ~1 pN at 6.5 nm s^{−1} (not shown), up to ~7 pN at 22 nm s^{−1} and up to ~25 pN at 88 nm s^{−1}.

To analyse quantitatively the unzipping of bridged areas, we converted our data to display contour length (see Supplementary Information). The steps shown in Fig. 2a probably reflect the disruption of individual H-NS-mediated bridges (note that occasional rebinding occurs—indicated by an asterisk). We determined the size of these steps using a step-fitting algorithm¹². When the sizes of the steps are plotted in a histogram (Fig. 2b) we see a distribution that can be fitted well to multiple gaussians with identical spacing, yielding a repeating unit of 3.6 ± 0.4 nm. It has not escaped our notice that this spacing equals the length of a helical repeat (~3.5 nm for B-DNA). The Fourier transform of the combined pairwise distribution function (pdf) of our fits shows distinct peaks (0.09, 0.14 and 0.29 nm^{−1}, corresponding to three, two and one helical turn(s) respectively), further strengthening our observation that H-NS binding occurs with variable spacing in register with the DNA helical repeat (Fig. 2b, insets; see Supplementary Fig. 2 and alternative analysis methods in Supplementary Figs 3 and 4). These analyses show that adjacent building blocks are positioned on opposite faces of two

¹Department of Physics and Astronomy and Laser Centre, Vrije Universiteit, De Boelelaan 1081, 1081 HV, Amsterdam, The Netherlands.

*These authors contributed equally to this work.

bridged DNA helices, as might be expected from DNA bridging models^{7,8}. Looking at the individual steps in a bridged region, we note that the density of H-NS binding is highly variable (Fig. 2c), which indicates that the H-NS units do not interact to achieve cooperativity⁸. Rather, apparent cooperative behaviour arises as an intrinsic property of DNA bridging proteins owing to duplex proximity⁷.

What are the structural implications of our observations for the layout of H-NS-bridged regions? Only structures of the isolated carboxy-terminal DNA binding domain (with a diameter of ~ 3 nm)¹³

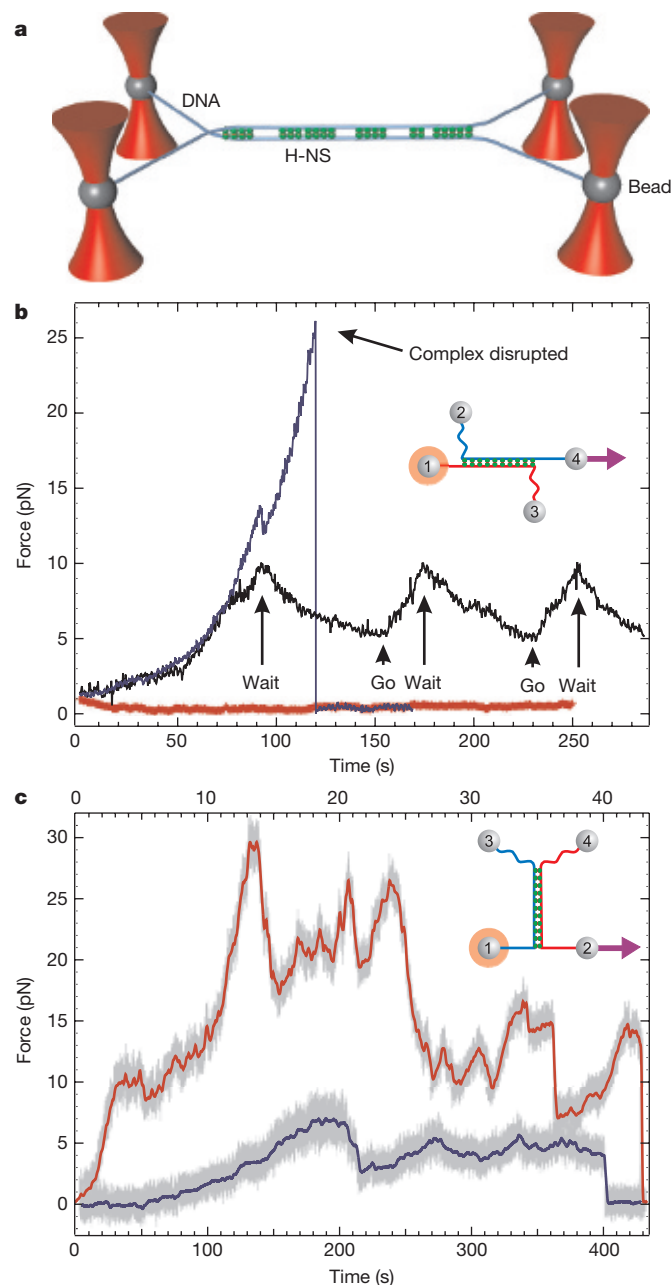


Figure 1 | Dual DNA manipulation. **a**, λ -DNA molecules linked by H-NS suspended between polystyrene beads held with optical tweezers. **b**, Shearing experiment. Beads are moved apart (at 22 nm s^{-1}) until the complex has been disrupted (blue) or bead movement is stopped at 10 pN and the complex is allowed to disassemble (black). DNA duplexes pre-coated with H-NS do not interact (red). Data displayed at 1 kHz . **c**, Unzipping experiment. Beads are moved apart until the DNA duplexes are separated (at 22 nm s^{-1} (blue) and 88 nm s^{-1} (red)). Data displayed at 2 kHz (grey) and 4 Hz (red/blue). Insets: schematic representations of the experiments.

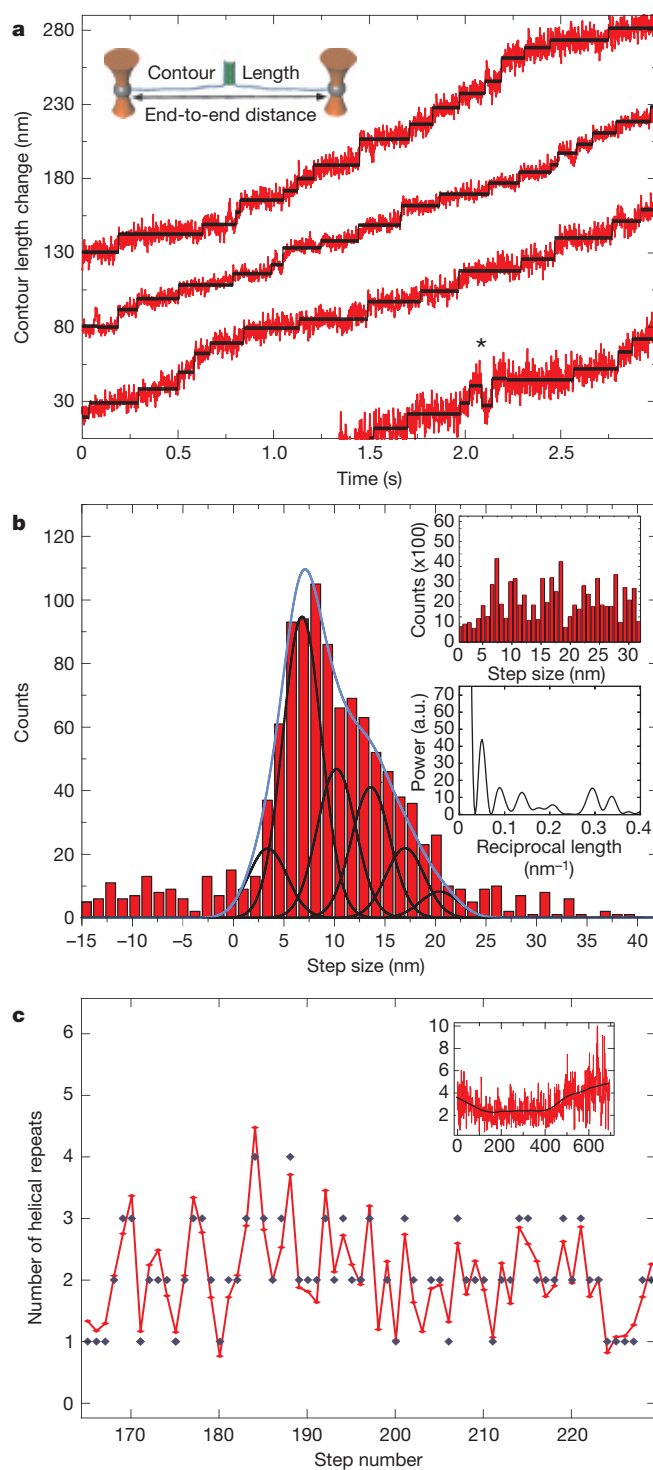


Figure 2 | Characteristics of H-NS bridges. **a**, One-sided contour length change as a function of time in unzipping experiments. Data displayed at 1 kHz (red) including staircase fits (black)¹². **b**, Histogram of all step sizes ($N = 904$). A multiple-gaussian fit (blue) with the individual gaussians (black). Inset, top: pdf of all the fitted steps for forces $> 10 \text{ pN}$ ($N = 458$). Inset, bottom: Fourier transform of the pdf. a.u., arbitrary units. **c**, Detailed view of the irregular spacing between adjacent H-NS bridges. Red squares correspond to a step sequence whose sequential order is plotted on the x-axis and the size of the step expressed in multiples of the DNA helical repeat on the y-axis. Blue squares, data rounded off to closest integer. Inset: Global view of spacing between H-NS bridges in a bridged region. Measured steps (red) and smoothed (locally-weighted-regression) data (black).

and the amino-terminal dimerization domain are available⁵. Two structures, which differ in the relative orientation of the dimerization interfaces (parallel⁸ or anti-parallel¹⁴) have been proposed (Fig. 3). About 10% of the steps can be assigned to an H-NS spacing of a single helical repeat (Fig. 2b). In combination with the absence of a regular (two-step) unbinding pattern, we can effectively rule out the tetrameric (or larger multimeric) assemblies of H-NS observed in solution^{15,16} as the building blocks of H-NS bridged regions. Therefore, in our model each building block is a dimer that effectively extends over approximately one helical turn on the DNA (Fig. 3). As steric hindrance by swivelling of the ~6-nm 'tail' (the dimerization domain)⁸ of the parallel H-NS dimer could yield the observed variations in the effective dimensions along the helix axis, it seems likely that the parallel H-NS structure prevails under our experimental conditions.

Little is known about the kinetics of the H-NS–DNA interaction. Usually, with dynamic force spectroscopy (DFS)¹⁷, interactions between bodies are explored by determining the rupture force/loading rate dependence at pre-set loading rates. In our approach, each individual experimental unzipping trace contains loading rate and unbinding force information for numerous individual rupture events in series, immediately yielding a dynamic force spectrum for the H-NS–DNA interaction (Fig. 4a). The resulting spectrum is characterized by two linear regions with different slopes, reflecting the presence of two well-defined kinetic barriers in the energy landscape¹⁷. These two barriers are located at distances $1.2 \pm 0.2 \text{ Å}$ ($x_{\beta 1}$) ($N = 34$) and $5 \pm 2 \text{ Å}$ ($x_{\beta 2}$) ($N = 479$) from the ground state along the direction of the applied force (see Supplementary Information). $x_{\beta 1}$ probably reflects the docking of the DNA binding domain to the DNA (presumably involving hydrogen bonds and Van der Waals interactions), whereas $x_{\beta 2}$ could correspond to a 'binding mode' that allows lateral diffusion¹⁸. When a diffusing H-NS dimer encounters a pre-existing bridge it becomes spatially restrained and will have ample time to cross the inner barrier. This would also make its binding to the second DNA duplex more likely and account for cooperativity. Therefore, an energy landscape with two well-defined barriers seems ideally suited to a protein that needs to bind to two DNA duplexes for functionality.

The thermal off-rate, k_{off} , of H-NS shows an exponential increase as a function of the applied force, F (Fig. 4b). Assuming an Arrhenius relation ($k_{\text{off}}(F) = k_{\text{off}}(0) \cdot e^{\frac{aF}{k_B T}}$; k_B is the Boltzmann constant; T is the temperature) yields a distance, $a = 2.0 \pm 0.3 \text{ Å}$, which is probably the same as barrier $x_{\beta 1}$ determined using DFS. The small shift in distance is probably a reflection of the underlying assumption that the barrier does not shift as a result of the application of force¹⁷. The high off-rate at zero external force ($k_{\text{off}} = 1.5 \pm 0.2 \text{ s}^{-1}$) indicates that there is substantial breathing of bridged regions, potentiating effective

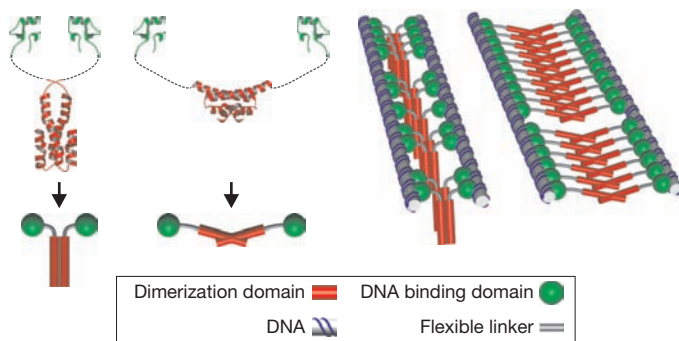


Figure 3 | Structural models of H-NS and H-NS–DNA₂ complexes. Left, the hypothetical dimer structures of full-length H-NS, as derived from the individually resolved structures of the dimerization^{8,14} and DNA-binding domains¹³. These domains are connected through an unstructured flexible linker. Right, models for H-NS–DNA interaction. H-NS interacts with two DNA duplexes as a dimer in either a parallel⁸ or anti-parallel¹⁴ configuration.

competition for DNA binding by other proteins *in vivo*. Assuming diffusion-controlled association with a typical rate constant of $k_{\text{on}} = 1.0 \times 10^5 \text{ M}^{-1} \text{ s}^{-1}$ (ref. 19) yields an equilibrium constant of $K_D = k_{\text{off}}/k_{\text{on}} = 1.5 \times 10^{-5} \text{ M}$, in good agreement with the K_D expected for sequence-unspecific binding²⁰. Using $\Delta G^\circ = RT \ln K_D$, where R is the gas constant, we estimate the Gibbs' free energy difference of complex formation for H-NS binding to be -27 kJ mol^{-1} , in line with bulk isothermal calorimetry estimates²¹. Our data therefore yield a conceptual energy landscape as shown in Fig. 4a (inset).

How do our structural and kinetic findings relate to an *in vivo* situation? H-NS evidently can bridge DNA duplexes and is abundant, which provides support for the idea that this protein could act as a topological insulator on DNA loops *in vivo*^{2,22,23}. H-NS binding nucleates in A/T-rich areas^{4,5}, which then probably recruit other areas without H-NS for binding *in trans*. This process is likely to be facilitated by plectonemic DNA supercoiling². In a number of cases, H-NS has been shown to have a functional regulatory role in which it acts as a roadblock at the stage of promoter clearance *in vitro*^{24–26} or during active elongation *in vivo*²⁷. How then can we explain that, in general, transcription *in vivo* seems not to be strongly affected by supposedly ubiquitous H-NS roadblocks? RNAP can exert forces up to 25 pN (ref. 28). Our data show that the maximum force for complex disruption is ~7 pN at an unzipping rate of 70 bp s^{-1} , indicating that at this rate the barrier can easily be overcome. The combination of cooperative binding and a high off-rate allows H-NS to stabilize DNA loops without compromising their susceptibility to dynamic reorganization (by transcription or by other NAPs^{2,29}).

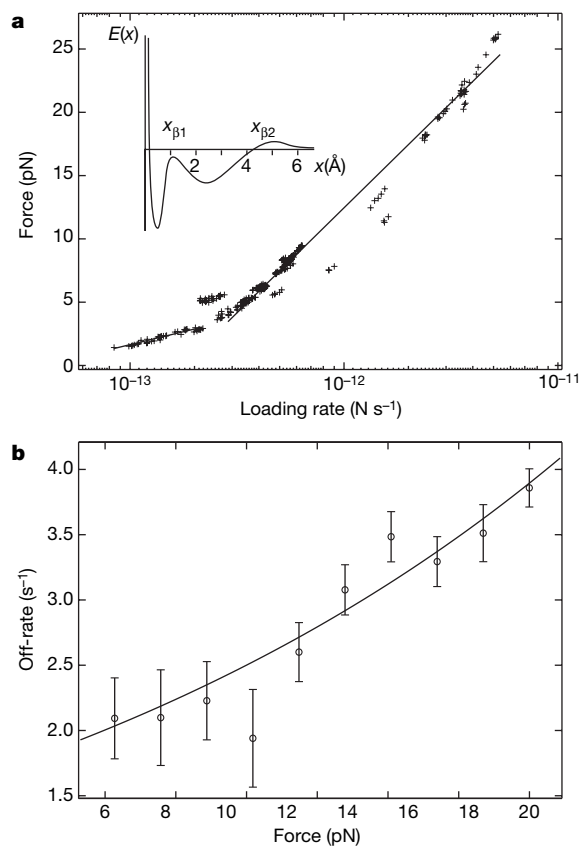


Figure 4 | Kinetics of the H-NS–DNA interaction. **a**, Dynamic force spectrum of the H-NS–DNA interaction. The two distinct regions in the spectrum have been fitted to straight lines. Inset: Conceptual energy landscape for the H-NS–DNA interaction. The x -axis shows the distance in Å. The y -axis displays the energy, E , in arbitrary units. **b**, The thermal off-rate of H-NS as a function of force with an Arrhenius fit ($N = 904$; error bars represent s.e.m.).

This is essential for *Escherichia coli* and other organisms in which restructuring the spatial organization of the genome is used for reprogramming of the global gene expression profile. Finally, these experiments demonstrate the potential of this dual DNA assay to study DNA bridging proteins and the ability to extract information that is accessible neither through bulk biochemistry nor with a single DNA approach. It also opens up a new range of single-molecule experiments including complex dynamic transactions between multiple DNA molecules.

METHODS

An optical-tweezers instrument was designed and constructed with which four micron-sized beads attached to the ends of two DNA molecules can be simultaneously trapped (Quadruple or Q-trap). Optical traps are generated using a 1,064-nm laser, which is first split on polarization. Subsequently one of the paths is led through an acousto-optic deflector to generate three time-shared traps at different locations. Experiments were performed in a flow cell with four channels. The use of laminar flow ensures that the channels do not mix³⁰. By moving the sample perpendicular to the channels we can efficiently catch, manipulate and incubate two single DNA molecules. Streptavidin-coated polystyrene beads are captured in the optical traps and subsequently incubated in the channel with biotinylated DNA solution. After DNA molecules have been captured between each set of beads they are tested for integrity and singularity in the buffer channel. The molecules are placed in the configuration required for the experiment and moved into the channel with H-NS solution. The position of the bead in the fixed trap is detected by back-focal-plane interferometry using a quadrant photodiode. End-to-end distance is obtained by determination of the distance between the stationary and the moving bead from video microscopy. A detailed description of experimental conditions, data analysis and dynamic force spectroscopy is provided in the Supplementary Information.

Received 8 May; accepted 26 September 2006.

- Schleif, R. DNA looping. *Annu. Rev. Biochem.* **61**, 199–223 (1992).
- Dame, R. T. The role of nucleoid-associated proteins in the organization and compaction of bacterial chromatin. *Mol. Microbiol.* **56**, 858–870 (2005).
- Johnson, R. C., Johnson, L. M., Schmidt, J. W. & Gardner, J. F. *Major Nucleoid Proteins in the Structure and Function of the Escherichia coli Chromosome* (ed. Higgins, N. P.) (ASM, Washington DC, 2005).
- Dorman, C. J. H-NS: a universal regulator for a dynamic genome. *Nature Rev. Microbiol.* **2**, 391–400 (2004).
- Rimsky, S. Structure of the histone-like protein H-NS and its role in regulation and genome superstructure. *Curr. Opin. Microbiol.* **7**, 109–114 (2004).
- Dame, R. T. *et al.* DNA bridging: a property shared among H-NS-like proteins. *J. Bacteriol.* **187**, 1845–1848 (2005).
- Dame, R. T., Wyman, C. & Goosen, N. H-NS mediated compaction of DNA visualised by atomic force microscopy. *Nucleic Acids Res.* **28**, 3504–3510 (2000).
- Esposito, D. *et al.* H-NS oligomerization domain structure reveals the mechanism for high order self-association of the intact protein. *J. Mol. Biol.* **324**, 841–850 (2002).
- Bustamante, C., Macosko, J. C. & Wuite, G. J. L. Grabbing the cat by the tail: manipulating molecules one by one. *Nature Rev. Mol. Cell Biol.* **1**, 130–136 (2000).
- Amit, R., Oppenheim, A. B. & Stavans, J. Increased bending rigidity of single DNA molecules by H-NS, a temperature and osmolarity sensor. *Biophys. J.* **84**, 2467–2473 (2003).
- Dame, R. T. & Wuite, G. J. L. On the role of H-NS in the organization of bacterial chromatin: from bulk to single molecules and back. *Biophys. J.* **85**, 4146–4148 (2003).
- Kerssemakers, J. W. *et al.* Assembly dynamics of microtubules at molecular resolution. *Nature* **442**, 709–712 (2006).
- Shindo, H. *et al.* Solution structure of the DNA binding domain of a nucleoid-associated protein, H-NS, from *Escherichia coli*. *FEBS Lett.* **360**, 125–131 (1995).
- Bloch, V. *et al.* The H-NS dimerization domain defines a new fold contributing to DNA recognition. *Nature Struct. Biol.* **10**, 212–218 (2003).
- Ceschini, S. *et al.* Multimeric self-assembly equilibria involving the histone-like protein H-NS. A thermodynamic study. *J. Biol. Chem.* **275**, 729–734 (2000).
- Smyth, C. P. *et al.* Oligomerization of the chromatin-structuring protein H-NS. *Mol. Microbiol.* **36**, 962–972 (2000).
- Evans, E. Probing the relation between force–lifetime–and chemistry in single molecular bonds. *Annu. Rev. Biophys. Biomol. Struct.* **30**, 105–128 (2001).
- Camacho, C. J. & Vajda, S. Protein docking along smooth association pathways. *Proc. Natl Acad. Sci. USA* **98**, 10636–10641 (2001).
- Eckel, R. *et al.* Single-molecule experiments in synthetic biology: an approach to the affinity ranking of DNA-binding peptides. *Angew. Chem. Int. Edn Engl.* **44**, 3921–3924 (2005).
- Jen-Jacobson, L. Protein–DNA recognition complexes: conservation of structure and binding energy in the transition state. *Biopolymers* **44**, 153–180 (1997).
- Ono, S. *et al.* H-NS is a part of a thermally controlled mechanism for bacterial gene regulation. *Biochem. J.* **15**, 203–213 (2005).
- Deng, S., Stein, R. A. & Higgins, N. P. Organization of supercoil domains and their reorganization by transcription. *Mol. Microbiol.* **57**, 1511–1521 (2005).
- Hardy, C. D. & Cozzarelli, N. R. A genetic selection for supercoiling mutants of *Escherichia coli* reveals proteins implicated in chromosome structure. *Mol. Microbiol.* **57**, 1636–1652 (2005).
- Dame, R. T., Wyman, C., Wurm, R., Wagner, R. & Goosen, N. Structural basis for H-NS-mediated trapping of RNA polymerase in the open initiation complex at the *rrnB* P1. *J. Biol. Chem.* **277**, 2146–2150 (2002).
- Schroder, O. & Wagner, R. The bacterial DNA-binding protein H-NS represses ribosomal RNA transcription by trapping RNA polymerase in the initiation complex. *J. Mol. Biol.* **298**, 737–748 (2000).
- Shin, M. *et al.* DNA looping-mediated repression by histone-like protein H-NS: specific requirement of E.σ⁷⁰ as a cofactor for looping. *Genes Dev.* **19**, 2388–2398 (2005).
- Dole, S., Nagarajavel, V. & Schnetz, K. The histone-like nucleoid structuring protein H-NS represses the *Escherichia coli* *bgl* operon downstream of the promoter. *Mol. Microbiol.* **52**, 589–600 (2004).
- Wang, M. D. *et al.* Force and velocity measured for single molecules of RNA polymerase. *Science* **282**, 902–907 (1998).
- Luijsterburg, M. S., Noom, M. C., Wuite, G. J. L. & Dame, R. T. The architectural role of nucleoid-associated proteins in the organization of bacterial chromatin: A molecular perspective. *J. Struct. Biol.* **165**, 262–272 (2006).
- van den Broek, B., Noom, M. C. & Wuite, G. J. L. DNA-tension dependence of restriction enzyme activity reveals mechanochemical properties of the reaction pathway. *Nucleic Acids Res.* **33**, 2676–2684 (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This research was supported by the Netherlands Organization for Scientific Research (NWO) through an NWO-Vernieuwingsimpuls grant (to G.J.L.W.), a FOM-projectruimte grant (to G.J.L.W.), an NWO-VENI grant (to R.T.D.) and a grant from ALW-NWO (to G.J.L.W.). We thank B. van den Broek, N. Goosen, J. van Mameren, E. Peterman, R. Wagner and C. Woldringh for their input and critical reading of the manuscript and J. Kerssemakers for providing the step-fitting algorithm and his assistance using it.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to G.J.L.W. (gwuite@nat.vu.nl). Correspondence may also be addressed to R.T.D. (rt dame@nat.vu.nl).

Structural basis for messenger RNA movement on the ribosome

Gulnara Yusupova^{1*}, Lasse Jenner^{1*}, Bernard Rees¹, Dino Moras¹ & Marat Yusupov¹

Translation initiation is a major determinant of the overall expression level of a gene^{1–3}. The translation of functionally active protein requires the messenger RNA to be positioned on the ribosome such that the start/initiation codon will be read first and in the correct frame. Little is known about the molecular basis for the interaction of mRNA with the ribosome at different states of translation. Recent crystal structures of the ribosomal subunits^{4–8}, the empty 70S ribosome⁹ and the 70S ribosome containing functional ligands^{10–13} have provided information about the general organization of the ribosome and its functional centres. Here we compare the X-ray structures of eight ribosome complexes modelling the translation initiation, post-initiation and elongation states. In the initiation and post-initiation complexes, the presence of the Shine–Dalgarno (SD) duplex causes strong anchoring of the 5'-end of mRNA onto the platform of the 30S subunit, with numerous interactions between mRNA and the ribosome. Conversely, the 5' end of the 'elongator' mRNA lacking SD interactions is flexible, suggesting a different exit path for mRNA during elongation. After the initiation of translation, but while an SD interaction is still present, mRNA moves in the 3'→5' direction with simultaneous clockwise rotation and lengthening of the SD duplex, bringing it into contact with ribosomal protein S2.

Thermus thermophilus 70S ribosome complexes corresponding to different functional states were crystallized and their structures were determined by X-ray analysis. Diffraction of some crystals extended to 3.9 Å resolution, and complete X-ray data were collected to 4.5 Å resolution (Supplementary Fig. S1 and Supplementary Tables S1 and S2). The improvement of crystal quality enables us to describe the complete atomic model for the path of an mRNA (SD(A)₄AUG(A)₉) in the initiation complex (Fig. 1a, and Supplementary Table S1). In a previous study, the mRNA had been localized at 7 Å resolution from Fourier difference maps, resulting in a gap of four nucleotides in the electron density at the position of the P codon and its 5'-flanking nucleotide (mRNA position –1 to +3)¹¹. The SD(A)₄AUG(A)₉ mRNA has the optimal distance of seven nucleotides between the core adenosine (–8) of the SD sequence¹⁴ (Fig. 1a) and the first nucleotide of the start codon (+1)^{15,16}.

Figure 1b shows the interaction between mRNA and transfer RNA in the 5.5-Å resolution structure of the 70S ribosome complex containing SD(U)₉UUU(U)₄ mRNA, designed without an AUG codon fixing the reading frame so that the ribosome has 'free choice' of the start codon. We found that the mRNA had moved and formed an extended and shifted SD duplex. Elongator tRNA^{Phe} was found in the A and P sites (A-tRNA and P-tRNA, respectively). This complex can be considered a good model for the post-initiation ribosomal state after the initiation of translation when SD interaction between mRNA and 16S RNA still exists and elongator tRNAs are already bound to the P and A codons. Cognate tRNA bound to the ribosomal

A site was observed previously at 7 Å resolution¹⁰, and non-cognate tRNA in the A site was found at 5.5 Å resolution in ribosome complexes containing a regulatory domain mRNA¹². In both cases, the A-site occupancy was low and the tRNA could be seen only in difference Fourier maps between the actual complex and a ribosome complex containing no mRNA. In the present post-initiation ribosome complex we found that tRNA binds in the A site with an occupancy close to that of P-tRNA and mRNA. The observed SD duplex has been extended to 12 nucleotides, although classical Watson–Crick base-pairing in the double helix occurs in only nine pairs (Fig. 1b). Thus, almost all nucleotides (from 1531 to 1542) of the single-stranded 3' end of 16S RNA are involved in the formation of the SD helix. The distance between the SD core adenosine and the start codon is also increased from seven (core A at position –8) to 12 (core A at position –13) nucleotides, and the length of the mRNA path between the 3' end of the extended SD helix and the P codon has increased from four to six nucleotides.

When the ribosome changes from the initiation to the post-initiation state, the mRNA 'relaxes', causing the SD duplex to lengthen, rotate and shift position. We observed this movement in difference Fourier maps between the post-initiation and initiation complexes (Fig. 1c). These maps show motion of the mRNA molecule in the 5'-end direction, an increase in the length of the SD duplex and the appearance of A-tRNA. The movement and extension of the SD lead to a much better resolved duplex, as a result of the increased number of interactions stabilizing the mRNA. The strands along the entire length of the double helix can be clearly distinguished in the electron density. Simultaneously, at negative contouring of the difference maps, we observed the disappearance of the SD duplex from its initiation complex position and the loss of density corresponding to the five 3'-terminal nucleotides (Fig. 1c). Translocation of the mRNA from the initiation to the post-initiation position is accompanied by a simultaneous clockwise rotation of the SD helix through about 70° (Fig. 1d). Although the overall ribosomal environment of the SD duplexes for the mRNA in initiation and post-initiation complexes is similar, the specific interactions differ because of the positional shift observed between the two (Fig. 1e, f, and Supplementary Information and Supplementary Fig. S2). In the initiation complex the SD duplex is in close contact with the amino terminus of ribosomal protein S18, whereas for the shifted and rotated SD duplex of the post-initiation complex the SD duplex now contacts protein S2 rather than S18.

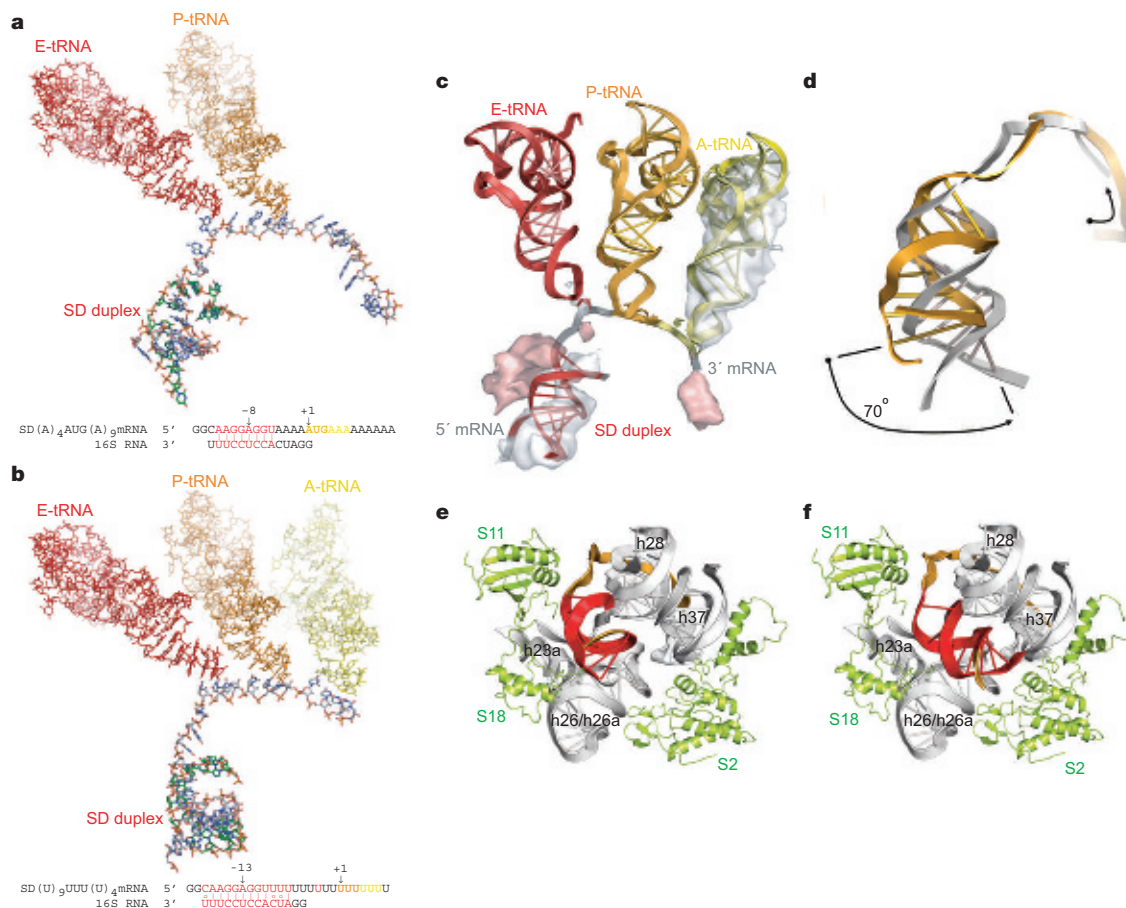
To study the mRNA structure upstream of the SD helix in the initiation complex, we designed an mRNA in which SD(A)₄AUG(A)₉ was extended at the 5' end by a poly(A) tail 33 nucleotides long (Supplementary Table S1). Initiation ribosome complexes were formed with the use of this (A)₃₃SD(A)₄AUG(A)₉ mRNA and initiator tRNA^{fMet}. Similarly, for the post-initiation complex,

¹Institut de Génétique et de Biologie Moléculaire et Cellulaire, 67404 Illkirch cedex, France.

*These authors contributed equally to this work.

To study the mRNA path on the 70S ribosome at the stage of elongation, we used a (U)₁₂ mRNA together with tRNA^{Phe} as a substrate for the A and P sites. Because of the low binding affinity of this elongator mRNA, which has neither an SD sequence nor an AUG codon, the occupancies of mRNA and A-tRNA^{Phe} were low, and these elements show up in difference Fourier maps very weakly (data not shown). However, some tRNA^{Phe} is clearly present in the A site, which confirms the presence of codon-anticodon interaction. To

The difference Fourier map between the 70S ribosome with (U)₁₂AUG(U)₉ mRNA, tRNA^{Phe} and tRNA^{fMet}, and the 70S ribosome with the nonanucleotide SD and tRNA^{fMet} shows continuous electron density for the (U)₁₂AUG(U)₉ mRNA and electron density for tRNA^{Phe} in the A site (Fig. 3c). This map clearly reveals that in the absence of the SD interaction, elongator mRNA (−12 to +12)



in red. **d**, Movement of mRNA from initiation (yellow) to post-initiation (grey) gives rise to a simultaneous positional rearrangement and clockwise rotation of the SD duplex. **e**, **f**, Detailed view of the ribosomal environment of the SD duplex in the initiation (**e**) and post-initiation (**f**) complexes.

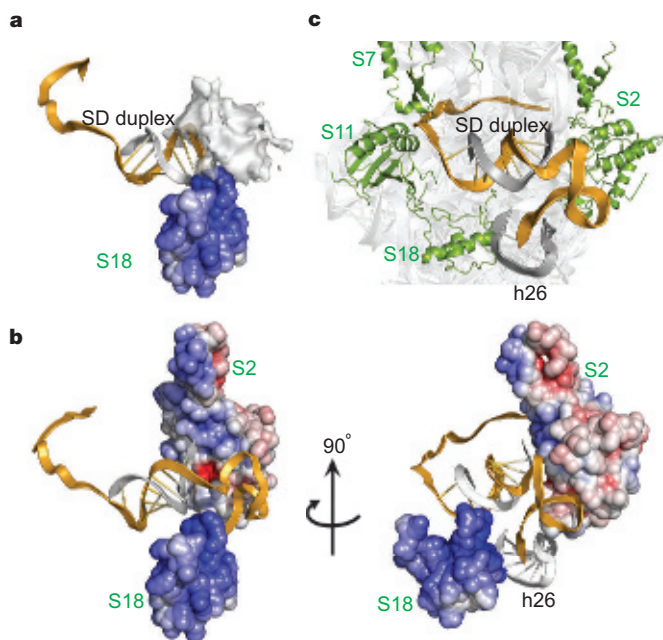


Figure 2 | Path of mRNA with a 5'-terminal extension through the ribosome. **a**, Interactions of (A)₃₃SD(A)₄AUG(A)₉ mRNA with protein S18 in the initiation complex. The globular density (grey, not modelled) may encompass the N-terminal part of protein S18 and about ten nucleotides of the poly(A) tail extension. **b**, Interactions of (U)₂₇SD(U)₉UUU(U)₁₀ mRNA with protein S2 in the post-initiation complex. **c**, Magnified fragment of the 30S subunit platform showing the environment of the SD duplex and the 5'-end extension of the post-initiation complex mRNA. A loop of the 5'-end extension of the mRNA enters a cavity in S2.

interacts with the ribosome from position -5 to position +12 only. A strong density is seen from position +12 to position -1, becoming progressively weaker from position -1 to position -5. The last seven nucleotides (-6 to -12) are not visible at all, implying that the 5' end of mRNA upstream of the E codon is flexible. Figure 3d, e shows the 70S ribosome in top view with all three tRNAs and elongation and post-initiation mRNAs, respectively. These are two 'snapshots' of the functional 70S ribosome: the first is at the post-initiation state (Fig. 3d), when SD interaction is still present in the mRNA but one or two codons have already been translated by the ribosome, and the second is at the elongation state, when all SD interaction is absent (Fig. 3e). In both cases the 3' end of mRNA wraps similarly around the 30S subunit.

Our observations suggest a simplified scheme for the movement of mRNA at different stages of translation (Fig. 3f). In the first step of initiation complex formation, the mRNA binds to the prokaryotic ribosome by establishing the SD duplex between the 5' end of the untranslated region and the accessible 3' end of the 16S RNA of the 30S ribosomal subunit. Subsequently, the SD duplex interacts with the 30S subunit platform and is oriented towards ribosomal protein S2 (presumably as seen in the ribosomal complex with SD nonanucleotide; Fig. 3b) and the rest of the mRNA molecule remains unbound (Fig. 3f, stage I). The following step is a simultaneous positional adjustment of the initiation codon into the ribosomal P site, where it is stabilized by codon-anticodon interaction with initiator tRNA^{fMet}, and a rearrangement of the mRNA SD duplex towards the ribosomal protein S18 (Fig. 1e, and Fig. 3f, stage II). At this stage of translation initiation, the mRNA is in a 'tense' conformation and is precisely positioned on the 30S subunit, so that the start codon will be read first and in the correct frame. Our findings of mRNA adjustment during the initiation process are in good agreement with earlier crosslinking experiments^{17–19}. Immediately after initiation, when one or several codons have already been translated by the ribosome but the SD interaction is still intact, a simultaneous

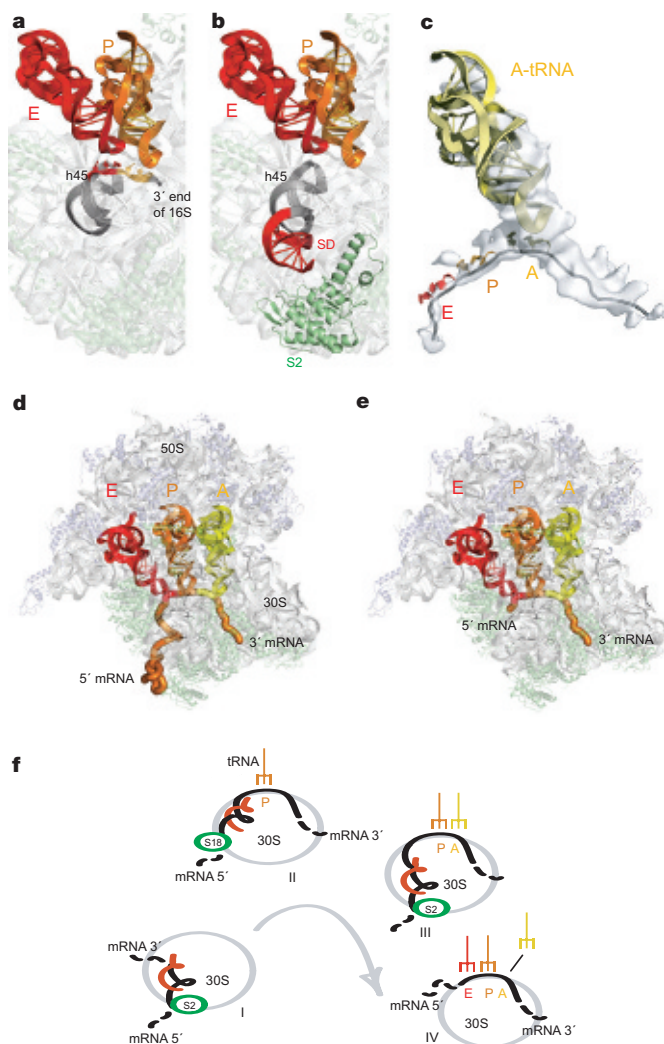


Figure 3 | mRNA path at the elongation step. **a**, Structure of the 3' end of 16S RNA (grey) in the ribosome complex without mRNA. **b**, SD nonanucleotide (red) in complex with the 70S ribosome and tRNA^{fMet}. The position of the SD duplex is similar to that seen in the post-initiation complex. **c**, Difference Fourier map between the ribosome complex containing elongator (U)₁₂AUG(U)₉ mRNA and the SD nonanucleotide. **d**, **e**, Top views of the 70S ribosome complex in the post-initiation state with mRNA containing a 5'-terminal extension (**d**) and in the elongation state (**e**). **f**, Simplified diagram of mRNA motion on the ribosome.

movement of the complete mRNA molecule in the 5' direction and a lengthening of the SD helix take place (Fig. 1f, and Fig. 3f, stage III). The SD helix is once again shifted towards ribosomal protein S2. Further translation of mRNA by the ribosome is accompanied by continuous rotation and movement in the 3'→5' direction and leads to melting of the SD interaction. Eventually, at some stage during elongation, the 5' end of mRNA will no longer interact at all with the ribosome (Fig. 3e, and Fig. 3f, stage IV) and will be accessible for other ribosomes, which can lead to the formation of a polysome.

METHODS

T. thermophilus 70S ribosomes were prepared and co-crystallized with purified *Escherichia coli* tRNA (tRNA^{fMet} and tRNA^{phe}) (Chemical Block) and mRNAs (Dharmacon), or without mRNA, using the conditions reported previously²⁰, with some modification in the ionic condition for complex formation (see Supplementary Methods). The mRNA samples were gel-purified before use in crystallization. Diffraction data (Supplementary Table S2) were collected at cryogenic temperatures at Swiss Light Source (SLS, Paul Scherrer Institut, Villigen, Switzerland). Data were integrated, processed and merged with the use of the HKL2000 package²¹. Refinement was done using the program

CNS²² with bulk solvent correction as described previously²³. Model building was performed in O (ref. 24), and figures were produced with the program PyMOL²⁵.

Received 27 June; accepted 27 September 2006.

Published online 18 October 2006.

- Gold, L. Posttranscriptional regulatory mechanisms in *Escherichia coli*. *Annu. Rev. Biochem.* **57**, 199–233 (1988).
- Draper, D. E. in *Escherichia coli and Salmonella: Cellular and Molecular Biology* 2nd edn (eds Neidhardt, F. C. et al.) 902–908 (ASM Press, Washington, DC, 1996).
- Kozak, M. Regulation of translation via mRNA structure in prokaryotes and eukaryotes. *Gene* **361**, 13–37 (2005).
- Wimberly, B. T. et al. Structure of the 30S ribosomal subunit. *Nature* **407**, 327–339 (2000).
- Ogle, J. M. et al. Recognition of cognate transfer RNA by the 30S ribosomal subunit. *Science* **292**, 897–902 (2001).
- Ban, N. et al. The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. *Science* **289**, 905–920 (2000).
- Schluzen, F. et al. Structure of functionally activated small ribosomal subunit at 3.3 Å resolution. *Cell* **102**, 615–623 (2000).
- Harms, J. et al. High resolution structure of the large ribosomal subunit from a mesophilic eubacterium. *Cell* **107**, 679–688 (2001).
- Schuwirth, B. S. et al. Structures of the bacterial ribosome at 3.5 Å resolution. *Science* **310**, 827–834 (2005).
- Yusupov, M. M. et al. Crystal structure of the ribosome at 5.5 Å resolution. *Science* **292**, 883–896 (2001).
- Yusupova, G. Z. et al. The path of messenger RNA through the ribosome. *Cell* **106**, 233–241 (2001).
- Jenner, L. et al. Translational operator of mRNA on the ribosome: how repressor proteins exclude ribosome binding. *Science* **308**, 120–123 (2005).
- Petry, S. et al. Crystal structures of the ribosome in complex with release factors RF1 and RF2 bound to a cognate stop codon. *Cell* **123**, 1255–1266 (2005).
- Schurr, T., Nadir, E. & Margalit, H. Identification and characterization of *E. coli* ribosomal binding sites by free energy computation. *Nucleic Acids Res.* **21**, 4019–4023 (1993).
- Ringquist, S. et al. Translation initiation in *Escherichia coli*: sequences within the ribosome-binding site. *Mol. Microbiol.* **6**, 1219–1229 (1992).
- Ma, J., Campbell, A. & Karlin, S. Correlations between Shine–Dalgarno sequences and gene features such as predicted expression levels and operon structures. *J. Bacteriol.* **184**, 5733–5745 (2002).
- La Teana, A., Gualerzi, C. O. & Brimacombe, R. From stand-by to decoding site. Adjustment of the mRNA on the 30S ribosomal subunit under the influence of the initiation factors. *RNA* **1**, 772–782 (1995).
- Rinke-Appel, J. et al. Contacts between 16S ribosomal RNA and mRNA, within the spacer region separating the AUG initiator codon and the Shine–Dalgarno sequence; a site-directed cross-linking study. *Nucleic Acids Res.* **22**, 3018–3025 (1994).
- Canonaco, M. A., Gualerzi, C. O. & Pon, C. L. Alternative occupancy of a dual ribosomal binding site by mRNA affected by translation initiation factors. *Eur. J. Biochem.* **182**, 501–506 (1989).
- Cate, J., Yusupov, M., Yusupova, G., Earnest, T. & Noller, H. X-ray crystal structures of 70S ribosome functional complexes. *Science* **285**, 2095–2104 (1999).
- Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
- Brünger, A. T. et al. Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
- Rees, B., Jenner, L. & Yusupov, M. Bulk-solvent correction in large macromolecular structures. *Acta Crystallogr. D* **61**, 1299–1301 (2005).
- Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).
- DeLano, W. L. *The PyMOL Molecular Graphics System* (DeLano Scientific, San Carlos, California, 2002).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank C. Schulze-Bries and the SLS staff for help with X-ray data collection. This work was funded by CNRS and INSERM.

Author Contributions G.Y. and L.J. contributed equally to this work. G.Y. performed the biochemical experiments and crystallized the ribosome complexes. L.J. solved the structures. All authors participated in discussing the results and writing the manuscript.

Author Information Coordinates and structure factors have been deposited in the Protein Data Bank under accession numbers 2HGR and 2HGU (initiation complex 30S and 50S subunits), 2HGP and 2HGQ (post-initiation complex) and 2HGI and 2HGJ (mRNA-free complex). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to M.Y. (marat@titus.u-strasbg.fr).

naturejobs

**THE CAREERS
MAGAZINE FOR
SCIENTISTS**

There are few easier ways for newspapers and magazines to generate a buzz than to print a list ranking the best of anything — from the greatest rock albums to the most essential books. And last month, *The Times Higher Education Supplement* followed suit, getting tongues wagging in academia around the world when it published its list of the 200 best universities.

The usual suspects made up the top 20, but the rankings awarded to the other 180 have proved a hot topic for conversation. This was abundantly clear to me while I was in Taiwan and Japan recently. Speaking to administrators there, it seemed that the most controversial point in the league table was the huge amount some institutions had moved in the space of 12 months. What did Beijing's Tsinghua University do, for example, to warrant it leaping from number 62 last year to 28 this year? And why did the French-speaking Brussels Free University plunge from 76 to 165 over the same period?

The answers lie in the indicators used as benchmarks to draw up the table. These include factors such as the number of citations per faculty member and the number of students and faculty members recruited from abroad. Although such indicators are useful, the administrators I spoke to in the Far East can't be alone in wondering whether these numbers can really tell anyone which university is better than another.

The people who may be most helped — or misled — by these scores are students and postdocs looking for the best place to begin their scientific career. One university administrator's son I spoke to, now weighing up which graduate school to attend, asked me about the importance of such lists. I told him that they help cement the reputations of world-class institutions; but I added that an institution's pedigree may not be the most important factor when choosing where to study. If you think in terms of tangibles such as funding, publications and available infrastructure, as well as intangibles such as granting first authorship on papers or teaching you new skills, then the quality of the principal investigator trumps prestige every time.

Paul Smaglik, *Naturejobs* editor

Editor: Paul Smaglik
Assistant Editor: Gene Russo

European Head Office, London
The Macmillan Building,
4 Crinan Street,
London N1 9XW, UK
Tel: +44 (0) 20 7843 4961
Fax: +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

European Sales Manager:
Andy Douglas (4975)
e-mail: a.douglas@nature.com
Business Development Manager:
Amelie Pequignot (4974)
e-mail: a.pequignot@nature.com
Natureevents:
Claudia Paulsen Young (+44 (0) 20 7014 4015)
e-mail: c.paulsenyoung@nature.com
France/Switzerland/Belgium:
Muriel Lestrange (4994)

UK/Ireland/Italy/RoW:
Loredana Milanese (4944)
Nils Moeller (4953)
Scandinavia/Spain/Portugal:
Evelina Rubio-Morgan (4973)
Germany/Austria/The Netherlands:
Reya Silao (4970)
Online Job Postings:
Matthew Ward (+44 (0) 20 7014 4059)

European Satellite Office
Germany: Patrick Phelan
Tel: +49 89 54 90 57 11
Fax: +49 89 54 90 57 20
e-mail: p.phelan@nature.com

Advertising Production Manager:
Stephen Russell
To send materials use London address above.
Tel: +44 (0) 20 7843 4816
Fax: +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

Naturejobs web development: Tom Hancock
Naturejobs online production:
Catherine Alexander

US Head Office, New York
75 Varick Street, 9th Floor,
New York, NY 10013-1917
Tel: +1 800 989 7718
Fax: +1 800 989 7103
e-mail: naturejobs@natureny.com

US Sales Manager: Peter Bless

Japan Head Office, Tokyo
Chiyoda Building, 2-37 Ichigayatamachi,
Shinjuku-ku, Tokyo 162-0843
Tel: +81 3 3267 8751
Fax: +81 3 3267 8746

Asia-Pacific Sales Manager:
Ayako Watanabe
e-mail: a.watanabe@natureasia.com



TREASURE IRELAND

Three years ago, marine biologists Kealan Doyle and Ken Maher found a lucrative niche: farming seahorses in Connemara, County Galway. The fish are sold to aquarium lovers and the medical markets in Hong Kong and Taiwan, where they are widely believed to be an aphrodisiac. Galway is Ireland's prime location for the ocean sciences, and a growing number of local researchers are involved in farming marine plants and animals. Doyle and Maher, who met in 1990 at the Galway–Mayo Institute of Technology, are part of a growing crowd of business-savvy scientists in Ireland who are reaping the benefits of globalization.

Picturesque Galway fulfils all the expectations of the tourists who flock to Ireland. But they are no longer the only arrivals. Talented young scientists, engineers, business managers and multinational companies are all heading to the country. Where it once suffered from mass emigration, Ireland now imports scientists and exports ideas.

Massive investment in science and education has been a key factor in Ireland's breathtaking economic rise. The most recent data from the European Union (EU) show that, from 1993 to 2004, the number of research and development (R&D) personnel in Ireland has nearly doubled to 16,000.

The scientific and economic boom began in Dublin and spread from there. In the counties of Galway, Limerick and Cork, the number of science students and research grants have been rising sharply, along with housing prices. Research income at University College Cork (UCC), for example, has risen from €21 million (US\$26 million) in 1997 to more than €62 million in 2005.

The Irish government is investing heavily in science and technology. As a result, career opportunities are becoming plentiful. Quirin Schiermeier reports.

Ireland's universities, such as Cork (above) and Galway (right), have their sights set on international renown.

All Irish science-funding agencies and government departments have now adopted quality benchmarks and peer-review standards similar to those applied by the US National Science Foundation. In the past, grants were reviewed at a national or local level. Now, they're reviewed with a view towards international R&D, meaning stricter review criteria and higher-quality reviewers.

The output and impact of Irish-based science has increased accordingly. The country is particularly strong in the agricultural sciences, in terms of number and impact of papers. But its citations performance also notably exceeds the average in molecular biology and genetics, immunology, geosciences, computer sciences and materials research.

"When I first came here in 1979 you were considered to be published if you had a letter in the *Irish Times*," says Michael Guiry, director of the Martin Ryan Institute, the marine research and teaching facility of the National University of Ireland, Galway (NUIG). "The progression since has been tremendous. If Ireland develops as well in the next ten years as it has in the past decade we will become a major scientific force."

Science funding and career opportunities improved dramatically



in the mid-1990s when the government began overhauling its low-wage, tourism-based economy. At the time, research was almost entirely funded by small grants from the EU's Framework programme for research. Under the Programme for Research in Third-Level Institutions, launched in 1998, some €600 million — matched by several hundred million from the US donor organization Atlantic Philanthropies — were injected, mainly into infrastructure and equipment. In addition, Science Foundation Ireland (SFI), a government-funded grant-giving agency set up in 2000, has invested some €650 million in research into biotechnology and information, and communications technology, seen as the main drivers of future economic growth.

Fresh approach

The National Development Plan for 2007 to 2013 foresees a further €3.8 billion for science, €2.7 billion of which will be spent in the next two years. Among other things, the government aims to double the number of PhD students to 8,000 by 2013, to meet industry's growing demand for a highly qualified workforce.

The government is attempting to restructure PhD training and improve graduates' skill sets as part of a recently launched initiative called Fourth-level Ireland. Targeted subjects include statistics, methods and ethics, but also basic business skills and the means to commercialize research. Starting next year, most PhD training courses will be extended to four years. "Career destinations are much greater than just academic, so we need to recast how we think about PhDs," says John O'Halloran, vice-head of UCC's college of science. "We really need to diversify opportunities, even though this will require a shift in the mindsets of supervisors and funding agencies."

The Irish strategy to strengthen domestic research capacities by attracting US and multinational industries has already borne fruit. Just last month, the soft drinks and snack food manufacturer PepsiCo announced plans to build a €10-million research facility in Cork, which will employ 25 chemists, food scientists and engineers. Leading computer and software companies, including Microsoft, IBM, Intel and Dell, are also producing ideas and technologies in Ireland. And the pharmaceutical industry centred on Cork is shifting from just manufacturing to R&D.

Expanding horizons

Meanwhile, Ireland's seven universities — small to medium-sized by European standards — have become more enterprise-minded, as have many individual researchers. SFI, through its Centres for Science, Engineering and Technology programme, supports campus-industry partnerships for up to ten years, with grants normally ranging from €1 million to €5 million per year. Additional money is available from the Industrial Development Agency, a government body that secures investment from overseas companies in Ireland. Enterprise Ireland, another government agency, helps and advises scientists who want to start a company or otherwise commercialize their research.

Two of SFI's seven new centres are on the rapidly growing NUIG campus: the Digital Enterprise Research Institute, a joint venture with Hewlett-Packard, and the Regenerative Medicine Institute, which investigates adult stem cells and gene therapy in partnership with



Bright future: (from top) Frank Barry, Fergus Shanahan and John O'Halloran are excited by Ireland's transformation so far.

medical devices manufacturers Medtronic and Charles River Laboratories.

In the past two years, says Nicholas Canny, the NUIG's vice-president for research, industry delegations have visited every two months to negotiate collaborations. "Public-private partnerships are the big push," he says. "We're a high-wage economy now, so we really do need to get the higher end of industry involved."

In August, the Alimentary Pharmabiotic Centre at UCC launched a €13.7-million project with GlaxoSmithKline aimed at finding new drug targets for gastrointestinal diseases. "This is true blue-skies research," says Fergus Shanahan, the centre's director. "We're not doing it for them — we're doing it with them."

More centres are to be created soon. In Galway alone, up to 100 new positions in the life sciences, from predoctoral to group-leader level, are likely to become available in the next two years, says Frank Barry, scientific director of the Regenerative Medicine Institute.

But Ireland's focus on industrial applications and economic growth has raised fears that curiosity-driven basic research might become neglected. "These concerns are real," says Guiry. The Irish science system has many parallels with EU research programmes, he notes, in that neither is very friendly to basic science.

"Commercial thinking is a good thing, as long as the mix is right and industry doesn't decide what's happening and what's not," says Barry. Still, training an ever-growing number of science students for jobs in industry makes some academics wonder how universities will cope with the intended doubling of PhD students. The government has no plan to foster a corresponding increase in the number of lecturers, senior scientists and supervisors, says Canny.

University challenge

The Irish research system doesn't support full-time researchers very well, says Canny, and many departments are already short of senior lecturers and experienced researchers. Some fear this could lead to a precarious imbalance. "You can't just flood the universities with students," says Guiry. "You have to give scientists the time to do the things they're supposed to."

A better-defined career structure for young scientists interested in pursuing an academic position is therefore high on the wish-list of Irish university staff. They seek a more predictable career structure, outlining a path to tenure-track positions and funding opportunities. Pensions, social security and incremental pay rises are also possibilities. As in the United States, postdocs have traditionally received few benefits. "A broadening of SFI's funding activities would help," says Guiry.

Meanwhile, Ireland is an increasingly popular destination for participants in the EU's Marie Curie programme, which provides two-year fellowships to graduate students and postdocs interested in doing research in another EU country. Since 2002, Ireland has brought in research worth €55 million from the programme, which is popular among both academic and industry researchers. Some 250 have come to Ireland, whereas only about 30 Irish scientists took up a Marie Curie position abroad.

Science in Ireland has become high-profile and very international in a remarkably short time. Researchers such as Guiry hope that a more consistent pattern of funding will guarantee that it stays that way. ■

Quirin Schiermeier is Nature's German correspondent.

Ireland funds great research

Maybe it's your turn.

Science Foundation Ireland, (SFI)

the national foundation for excellence in scientific research is investing in academic researchers and research teams who are most likely to generate new knowledge, leading edge technologies, and competitive enterprises.

SFI has a flexible grants and awards portfolio and several times a year issues calls for proposals from scientists and engineers. SFI also accepts unsolicited proposals throughout the year.

Principal Investigator Programme for outstanding researchers, normally ranging between €50,000 - €250,000 per year for a three to four year period.

Centres for Science, Engineering and Technology (CSET) linking academic and industrial researchers together on high-end research programmes, grants ranging up to €20 million over five years.

Strategic Research Clusters (SRCs) will create clusters of internationally-competitive researchers from academia and industry, particularly Irish-based industry. Grants will be for three years normally range from €500,000 to €1.5 million per year.

Research Professor Recruitment Awards for outstanding researchers, with particularly distinguished international reputations, awards normally ranging up to €500,000 per annum for up to two years.

Research Frontiers Programme (RFP) supporting all areas of high-quality, novel exploratory research in the third-level sector, normally ranging up to €150,000 over three years.

E.T.S. Walton Visitor Awards supporting leading international scientists who visit Ireland to undertake research for up to one year, normally ranging up to €200,000.

President of Ireland Young Research Awards (PIYRA) attracting to Ireland and supporting Irish researchers within five years of completing their Ph.D., normally up to €1 million over five years.

Undergraduate Research Experience & Knowledge Award (UREKA) supports active research participation by undergraduate students during the Summer.

Secondary Teacher Assistant Researchers (STARs) supports second-level teachers in conducting research in Irish laboratories during summers.

Workshops and Conference Grants supporting significant scientific meetings taking place in Ireland, involving Irish scientists and research bodies.

Apply for an SFI award or learn more about our programmes at www.sfi.ie



Built for Research

Science Foundation Ireland
Wilton Park House, Wilton Place, Dublin 2, Ireland
tel +353 1 607 3200 fax +353 1 607 3201
email info@sfi.ie www.sfi.ie

www.sfi.ie

SP90393R



Environmental Change Institute
National University of Ireland, Galway, Ireland



National University of Ireland, Galway
Ollscoil na hÉireann, Gaillimh



MARIE CURIE ACTIONS

National University of Ireland, Galway

EXPERIENCED RESEARCHERS

Project: Transfer of Knowledge required to understand cross-talk between host and pathogen during gastrointestinal infections in humans (GAMIDI)

Applications are invited for five experienced researcher (ER) posts available at the National University of Ireland (NUI), Galway, Ireland resulting from the award of an EU 6th Framework Marie Curie Transfer of Knowledge Grant - Transfer of knowledge required to understand cross-talk between host and pathogen during gastrointestinal infections in humans (GAMIDI). The Human Pathogen Cluster, located at the Department of Microbiology and facilitated by the Environmental Change Institute, will conduct research in the usage of advanced technologies to investigate the interaction between the human host and key gastrointestinal bacterial pathogens.

GAMIDI-ER1: Signalling events during the inflammation response triggered by *Vibrio parahaemolyticus*. Preference will be given to candidates with expertise in eukaryotic cell signalling analysis, transfection and heterologous protein expression in eukaryotic cells and/or siRNA methodologies. (Principal Investigator: Dr. Aoife Boyd)

GAMIDI-ER2: Molecular dissection of stress-induced signal transduction in *Listeria monocytogenes*. This researcher will be expected to have a strong background in protein biochemistry and molecular biology and have a keen interest in bacterial regulatory networks. (Principal Investigator: Dr. Conor O'Byrne)

GAMIDI-ER3: n-Gradostat culture to model gastrointestinal tract population dynamics *in vitro*. Candidates should be experienced in complex chemostat systems, preferably with experience of complex mathematical modelling of *in vitro*-generated bacterial population data. (Principal Investigator: Dr. Gerard Fleming)

GAMIDI-ER4: Examination of glycosyltransferase gene expression in the pathogenesis of *Helicobacter pylori* and *Campylobacter jejuni*. The researcher should have experience in transcriptomics and possess an ability to develop and potentially exploit gene microarray technology in the study of bacterial pathogens. (Principal Investigator: Prof. Anthony Moran)

GAMIDI-ER5: Proteomic characterisation of the adaptive acid tolerance response of *Campylobacter jejuni*. Preference will be given to candidates with a background in prokaryotic cell protein analysis methodologies, with emphasis on the application of proteomics. (Principal Investigator: Dr. Cyril Carroll)

The positions are tenable for 24 months and salaries are based on the Marie Curie ER scales. Candidates must be eligible under FP6 Marie Curie Fellowship regulations. The deadline for applications is November 30th, 2006.

Please send CV and cover letter with names of three referees quoting the job reference to: Caroline Murphy, Environmental Change Institute, NUI, Galway, Ireland. Tel: +353-91-493548 (caroline.murphy@nuigalway.ie)

SP90676R

Microbiology Department Trinity College Dublin

The Department of Microbiology within the Vice-Deanery of Genetics and Microbiology (www.tcd.ie/Microbiology/index.php) invites applications from researchers who are capable of winning personal external support from national or international funding agencies (such as Science Foundation Ireland and the Wellcome Trust) to establish leading-edge research programmes in Systems Microbiology and/or Infection Biology.

In the first instance, applicants should contact Professor Charles Dorman (cjdorman@tcd.ie) or Professor David J McConnell (dmcennll@tcd.ie) for informal discussions.



The University of Dublin
Trinity College



SP91125



Career articles from Naturejobs

- **Naturejobs Prospect:** quick takes on career implications of current events
- **Naturejobs Special Report:** examinations of jobs issues on both sides of the bench
- **Naturejobs Careers & Recruitment:** discipline-by-discipline exploration of opportunities
- **Naturejobs Regions:** tours of scientific hubs
- **Naturejobs Movers:** traffic reports that follow high-profile scientific globe-trotters and sector-hoppers
- **Naturejobs Postdocs & Students:** guides to taking a step toward a permanent position

www.naturejobs.com

naturejobs
making science work

nature publishing group

Need to find
the ideal
candidate
fast?

Visit
**www.
naturejobs
.com**

to discover
how applicants
can respond
directly to you
by email.

naturejobs
making science work



TRINITY COLLEGE DUBLIN

The University of Dublin



www.tcd.ie/vacancies

Trinity College is recognised internationally as Ireland's premier university and is the only university to rank in the top 100 world universities (78th) and amongst the top 50 European universities (25th). We are recruiting world class leaders in research and education to advance our research strengths, develop our fourth level graduate education and build on our excellence in third level undergraduate education.

Our strategic priorities of research and education are also aligned with contribution to the national goal of fostering Ireland's cultural and economic vibrancy.

Smurfit Institute of Genetics, Trinity College Dublin
and

Adelaide and Meath Hospital, incorporating
The National Children's Hospital (AMNCH) Ireland



**THE ADELAIDE & MEATH
HOSPITAL, DUBLIN**
INCORPORATING
THE NATIONAL CHILDREN'S HOSPITAL

Lectureship in Medical Molecular Genetics

The Smurfit Institute of Genetics and the Adelaide and Meath Hospital, inc. The National Children's Hospital (AMNCH) invite applications for a newly established permanent lectureship to be filled in 2007.

The staff of the Smurfit Institute have a wide and sustained commitment to high quality teaching and research. AMNCH is one of the two major teaching hospitals of Trinity College Dublin. The new lecturer will be expected to develop his/her own research programme in association with one or more clinical teams at the Hospital, through competitive funding, for example from Science Foundation Ireland, The Wellcome Trust, Health Research Board etc.

Applications will be welcomed from people working in any field of Molecular Medical Genetics.

Appointments will be made on the scale €34,678 – €69,985/€72,317 - €79,489 per annum. Appointment will accord with qualification and experience to date. Further information may be obtained from:

Professor David McConnell, Vice-Dean, Smurfit Institute of Genetics, Trinity College, Dublin 2, Ireland.
Email: david.mcconnell@tcd.ie

Applications with the names of three referees should be received by 31 January 2007 and addressed to:
Christine Devlin, Recruitment Executive, Staff Office, Trinity College, Dublin 2, Ireland.
Email: cdevlin@tcd.ie / Telephone +353 1 896 8489

Trinity College is an Equal Opportunities Employer

SP91295

The Office of Science, Department of Energy is seeking a motivated and highly qualified individual to serve as the

Associate Director

Office of Biological and Environmental Research

As such, you will provide leadership and direction in establishing vision, strategic plans, goals, and objectives for the research activities supported. You may apply through two different methods - one is for a

Senior Executive Service appointment and the second is for an Intergovernmental Personnel Act appointment. The announcement number is SES-SC-HQ-005. The announcement opens on November 6 and closes on December 21.

Visit <http://www.usajobs.opm.gov/>
for more information and for instructions concerning application procedures

NW90953R



THE SCHOOL OF PHARMACY

University of London

Department of Pharmaceutical &
Biological Chemistry

Postdoctoral Research Fellow

We have a vacancy for a Research Fellow in Biomimetic Diversity Oriented Synthesis. Applicants should hold a PhD in synthetic (natural product) chemistry. Experience in solid phase synthesis and structure elucidation of complex natural products and knowledge of antibacterial/anticancer natural products are desirable. The post is available for two years.

The salary range will be from £26,892 - £30,779 including London Allowance (effective from 1st February 2007).

Informal enquiries should be made to Dr. John Moses (tel. 020 7753 5804, email: john.moses@pharmacy.ac.uk).

For application details for this position, please view the School's website <http://www.pharmacy.ac.uk/jobs.html> email personnel@pharmacy.ac.uk, or telephone the Personnel Office on +44 (0) 20 7753 5813. Please quote reference RF-BC(JM).

Closing Date: Friday 22nd December 2006.

The School of Pharmacy is committed to equality of opportunity and selection on merit

91338R

Visit
www.naturejobs.com

to seriously improve
your career prospects.

naturejobs
making science work

MOVERS

John McNeil, scientific director, Program for Appropriate Technology in Health, Malaria Vaccine Initiative, Bethesda, Maryland



2006: Chief, Laboratory of Advanced Clinical Development, Dale and Betty Bumpers Vaccine Research Center, National Institute of Allergy and Infectious Disease, Bethesda, Maryland

2001-03: Deputy director, Division of Retrovirology, Walter Reed Army Institute of Research, Washington DC

John McNeil was certain of two things early on in his career: he would become a doctor and he would not follow in his parents' military-service footsteps. But burdened with debt after medical school at Wake Forest University in Winston-Salem, North Carolina, and Harvard School of Public Health, he reluctantly broke his pledge and joined the army for a health-profession scholarship. It was the start of a 23-year military career at the cutting edge of HIV vaccinology.

While finishing his residency at Walter Reed Army Hospital, McNeil also pursued an MS in public health at Harvard, and turned his focus to vaccines. He was among the first to use diagnostic tests for HIV infection. "My interest in infectious diseases and public health got swept up in AIDS," he says of the early days of the epidemic. He spent the late 1980s documenting the frequency and distribution of HIV infection in new populations. In the 1990s, he identified candidate HIV vaccines to test in Thailand — an ideal location because of its low genetic diversity. McNeil retired from the army in 2003, a day after the start of a 16,000-person phase III efficacy trial for one of his candidate vaccines. He decided his army work was complete; the best opportunities were elsewhere.

So McNeil moved to the vaccine research centre at the US National Institutes of Health. There, he shepherded candidate vaccines into larger field trials and explored new approaches such as DNA vaccines. Lamenting the limited cross-fertilization of ideas among experts in HIV, malaria and tuberculosis, McNeil became increasingly interested in applying his knowledge of HIV to malaria. Recent work on inducing cellular immunity to HIV should be more vigorously applied to malaria, he suggests.

Jim Tartaglia, vice-president for R&D at Sanofi Pasteur's facility in Canada, recommended McNeil for his latest post, as scientific director of the Program for Appropriate Technology in Health's Malaria Vaccine Initiative. He lauds McNeil's experience with novel vaccine-development technologies and his ability to partner industry, academia and non-governmental organizations, a crucial skill for developing vaccines in multiple international locations.

McNeil acknowledges the many challenge ahead: for example, developing live virus vector approaches to induce immunity early on in malaria's complex life cycle. Nevertheless, he remains optimistic, and predicts that there will be an effective malaria vaccine within the decade.

Virginia Gewin

BRICKS & MORTAR

A step between bench and bedside

A new clinical research facility at the National Institute of Environmental Health Sciences (NIEHS) in Research Triangle Park, North Carolina, aims to narrow the gap between research and patient care, and investigate how environmental exposures affect health. For the first time, scientists at the NIEHS, part of the US National Institutes of Health (NIH), will have on-campus clinical-research space.

The move will help the NIEHS embrace translational research, a critical NIH roadmap component that transforms bench science into medical practice, says William Martin, director of translational research at the NIEHS. Having an onsite clinical research unit is a necessary step towards turning basic research findings into patient care.

Initial topics of study will include the environmental triggers and inflammatory mechanisms of airway diseases such as asthma and chronic obstructive pulmonary disease. NIEHS director David Schwartz says that the centre will help scientists "move the field more towards human health than it has been in years, and contribute to an understanding of human biology".

The 1,100-square-metre unit will be built close to the main NIEHS building with room for up to 40 study subjects

daily. It is expected to begin accepting out-patients by summer 2007, and Schwartz hopes it will eventually take in-patients too.

The unit will provide routine evaluations, biological-sample collection and processing, pulmonary-function testing and bronchial-sampling capabilities. Exposure chambers will allow researchers to determine responses to different types of air pollution, such as ozone, in healthy people, those with chronic respiratory diseases, and those with genetic polymorphisms that may influence their reaction.

Future research interests include reproductive, neurodegenerative and metabolic diseases. Schwartz also aims to develop sophisticated imaging techniques for the lungs, bones, brain and metabolic processes.

The NIEHS will invest heavily in teaching principles of environmental-health science to physician-scientists at the unit, where Schwartz aims to increase their number from 6 to 12. He hopes the blend of basic research and clinical facilities will encourage PhD investigators to participate in clinical activities derived from their research. There will also be 20 technicians, nurses and physician assistants to run the unit and its research projects.

Hannah Hoag

GRADUATE JOURNAL

Done deal

I took the stand to defend my dissertation research three weeks ago. I have a hard time remembering what actually took place on that day, but I know that I passed. I do remember that the weather was beautiful, although it was exceptionally hot, even for Hawaii. I replaced my normal work attire — a pair of surf shorts and a T-shirt — with slacks and an aloha shirt, the closest I've come to wearing a jacket and tie during my nine years in Hawaii.

I wasn't all that nervous because I had difficulty contemplating that I was really about to defend. After practising my talk aloud to myself in front of a mirror a final time, which always makes me feel like an idiot, I felt the need to review the literature and my old notes. So with two hours to go, in an absolutely random fashion, I started to skim through a number of papers and notebooks in order to make up for anything I may have missed during the past five years. Remarkably, I actually got some questions on the topics I was able to cover. The defence itself went smoothly and, according to the 'objective' opinion of my mother, it was very good — although I am not sure how much of it she understood.

Despite the three weeks that have passed, I still cannot grasp that it actually happened. However, I am happy to be back wearing a T-shirt and shorts again.

Andreas Andersson is a final-year PhD student in oceanography at the University of Hawaii.

An unfortunate book tour

Who wants to live for ever?

Will Heydt-Minor

Johnny Morton stood in the trapezoid silhouette of the New York Public Library. Four marble statues of literary figures guarded the entryway under an awning propped up by gleaming black columns. Clarke, Wells, Verne, Bulgakov, all gazing westward across the plaza and beyond. Above this was the building proper, a central stalk flanked by sloping supports. At the summit was a statue of Calliope sprawled on a golden litter.

Johnny watched the sun tumble from its noontime zenith and elongate the library's shadow, framing the muse in flaming scarlet.

He checked his watch. It was time to enter, but he couldn't bear to move. The library was simply too beautiful. Finally, he took a deep breath and walked through the bronze doors.

Inside, hundreds of people mostly older than him huddled in small groups. The interplay of hushed voices sounded like light rain on grass.

He looked up and was overcome by vertigo. The building's exterior had belied its true size. A hundred circular levels shot upwards, the hollow middle culminating in a faint green ceiling.

Suddenly a man's voice echoed in the main room.

"Welcome to this once-in-a-lifetime occasion!"

Immediately everyone fell silent. Johnny stood on tiptoe to see who the speaker was, but the people in front were too tall.

The voice continued: "I can think of nothing better to commemorate this new library than a rare appearance by our revered man of letters, Dr Heinrich Kravitz!"

The crowd cheered enthusiastically.

"Unfortunately," the voice added, "I can allow only a handful in at one time. Will the closest 25 please approach?"

Johnny lowered his head and shoved through the crowd, ignoring hateful looks and offended exhalations. A few moments later he emerged in front, just in time to join the first group.

As they walked down a hallway, he glimpsed the speaker. Curiously, the man wore a lab coat with a name-tag that read Boreal-PharmaCo.

The group arrived at a large circular room with green carpeting. The ceiling

was capped by a glass dome, and in the centre, ringed by sunbeams, was a low dais. Johnny felt like he was inside an emerald.

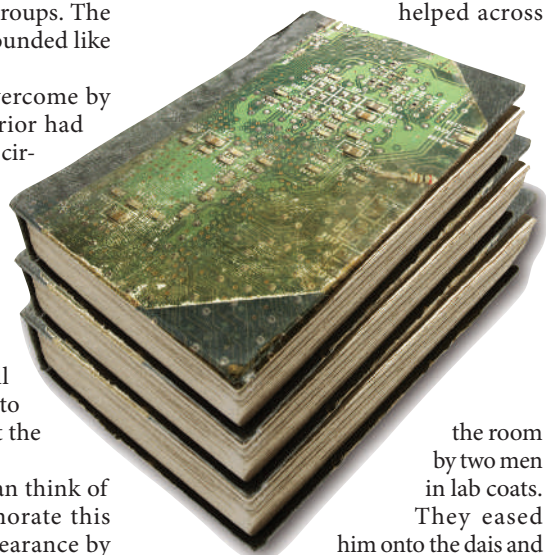
"Have a seat my friends," their guide said softly. "Dr Kravitz will be in shortly."

Johnny found a spot near the back of the group. After a few minutes, anxious whispers began to spread through their ranks. Johnny overheard one older boy say to another, "This is his first public appearance in 40 years — maybe he's scared?"

"Maybe he's imaginary," the other replied.

Johnny snorted. They didn't know the real Heinrich Kravitz. The good doctor would show, he was sure of it.

The whispering became fully fledged conversation, then grumbling and bickering. Finally, just as someone rose to leave, the doors swung open and three men entered. The group collectively took a breath and stood up. Johnny climbed a bookshelf to get a better view. He saw Dr Kravitz being helped across



the room by two men in lab coats. They eased him onto the dais and retreated.

"Good day, my children," the doctor said, and the entire group exhaled and sat down. He smiled broadly. "I have written a new book about my experiences as the longest-lived man in human history, entitled *Human and Trans-human*. I will now read selected passages."

He cleared his throat and tilted his head upwards. "Chapter One: the thermoplast heart is a device I postulated in my early work *Alpha-Centauri or Bust!* With the help of Boreal-PharmaCo it was invented two decades ago and placed inside me. In five years it will hit the market, readily available to all daring consumers. Chapter Two: the xenograft-reservoir is a device I

postulated in my early work *The Corona of Eternity*. With the help of..."

"Wait!" Johnny called.

The entire group turned: the men in lab coats exchanged glances.

"Dr Heinrich, I have a question."

The doctor looked down and smiled. "Please, call me Henry."

"Henry, wasn't your xenograft-reservoir actually in the short story *Marooned!*?"

The doctor was silent for a time. Then he spoke in an emotionless voice. "We hauled the body from the wreckage. Jim winced when he recognized our pilot, Smitty. 'Dammit if the old fool didn't give his all,' he whispered. I said a prayer and detached his xenograft reservoir."

"Exactly!" Johnny cried. "That's from *Marooned!*, the first story you co-wrote with Margareta."

The doctor stared blankly. "Who is Margareta?"

Johnny was dumbfounded. "Why, your wife..."

"My... wife?" he said, mispronouncing the word. "My wife? My wife?"

The two men in lab coats rushed forward and seized the doctor. One of them raised his hands. "The reading is over. Dr Kravitz is ill and needs our assistance."

The group looked around uncertainly.

"Leave!" he shouted.

They shrugged and filed out the exit. But Johnny hid behind a bookshelf and watched.

"I knew this stupid thing wasn't ready for the road," the first man said.

"How else were we supposed to de-bug him?" said the other.

The first detached Dr Kravitz's pate, revealing a wet grey pouch. He stuck his index finger in and swirled it around. "I knew it, that damn kid fried his back-up memory banks. This'll take months to fix."

They grabbed Dr Kravitz by the arms and unceremoniously dragged him from the room.

Meanwhile, Johnny remained hidden. Deep within he felt an odd sensation, a sort of emptiness, as if something had been forcibly removed. He buried his head in his arms and swallowed a sob.

Heinrich Kravitz was dead.

Will Heydt-Minor makes his home in Colorado. He writes for various Denver-based literary magazines and, together with Evan Cotageorge, publishes the science-fiction comic *A Day in the Life of Theodore Welkin*. He is currently seeking a degree in linguistics.

JACEY